

METRON

RIVISTA INTERNAZIONALE DI STATISTICA — REVUE INTERNATIONALE DE STATISTIQUE
INTERNATIONAL REVIEW OF STATISTICS — INTERNATIONALE STATISTISCHE ZEITSCHRIFT

DIRETTORE PROPRIETARIO — DIRECTEUR ET PROPRIÉTAIRE
EDITOR AND PROPRIETOR — HERAUSGEBER UND EIGENTHÜMER

Prof. Dott. Corrado Gini, *direttore dell'Istituto di Statistica e Politica Economica della R. Università di Roma, presidente dell'Istituto Centrale di Statistica del Regno d'Italia*

COMITATO DIRETTIVO — COMITÉ DE DIRECTION
EDITORIAL COMMITTEE — DIREKTION-KOMITEE

Prof. A. Andréadès (Athènes) — **Prof. A. E. Bunge** (Buenos Ayres) — **Prof. F. P. Cantelli** (Roma)
Prof. C. V. L. Charlier (Lund) — **Prof. F. v. Fellner** (Budapest) — **Prof. A. Flores de Lemus** (Madria)
Prof. M. Greenwood (London) — **Sir G. H. Knibbs** (Melbourne) — **Ing. L. March** (Paris)
Prof. A. W. Methorst (La Haye) — **Prof. A. Julin** (Bruxelles) — **Prof. R. Pearl** (Baltimore)
Prof. H. Westergaard (Copenhagen)

AMMINISTRATORE — ADMINISTRATEUR — MANAGER — VERWALTER

Dott. Silvio Orlandi, *Istituto di Statistica e Politica Economica della R. Università di Roma.*

SEGRETARI DI REDAZIONE — SECRÉTAIRES DE RÉDACTION
EDITORIAL SECRETARIES — REDACTIONSSECRÉTAIRE

Prof. Luigi Galvani — **Dott. Mario Saibante**

CECIL CALVERT CRAIG

An Application of Thiele's Semi-invariants to the Sampling Problem

ROMA

AMMINISTRAZIONE DEL « METRON »
R. UNIVERSITÀ — ISTITUTO DI STATISTICA
:: :: E POLITICA ECONOMICA :: ::

Ann Arbor 1928.



CECIL CALVERT CRAIG

**An Application of Thiele's Semi-invariants
to the Sampling Problem**

Introduction

The late Professor A. A. TSCHUPROW in a series of papers ⁽¹⁾ has made a sustained and successful effort to obtain detailed and exact results in the problem of sampling for statistical characteristics from an infinite parent population. But to push his methods further and obtain more of these results, which are certainly needed, causes any would-be investigator to pause and inquire into the possibility of devising a still more powerful method, especially one that would simplify and reduce the bulk of the algebra involved.

As a more powerful and at the same time more simple method than that used by TSCHUPROW I offer in this paper one based on the theory of the semi-invariants of THIELE ⁽²⁾. I shall show how exact results can be obtained by a process of mechanical computation which can be shown to and carried out by any reasonably intelligent undergraduate assistant. I shall develop the method for the distributions of moments about the mean and for the distributions of two moments about the mean. It will be shown how particular terms of a result can be computed separately, and how results to any desired degree of approximation may be obtained with an accurate measure of the magnitude of the terms omitted. Examples of the calculations and

⁽¹⁾ TSCHUPROW, A. A., *On the Mathematical Expectation of Moments of Frequency Distributions*; « *Biometrika* », Vol. XII (1918-1919), pp. 140-211, and Vol. XIII (1920-1921), pp. 283-309.

⁽²⁾ THIELE, T. N., *Theory of Observations*; Charles and Edwin Layton, London, 1903; Chapters VI-VIII.

new results will be given. Then by taking advantage of a peculiar and important property of semi-invariants and with the aid of the semi-invariants of the distribution function for the second and the third moments about the mean and of the one for the second and fourth moments about the mean, I shall take up the problem of the distribution of the relative or absolute moments, α_3 and α_4 , defined in the usual way, with considerably better success than has hitherto been won.

CHAPTER I

Semi-invariants and Moments

I shall begin with a brief discussion of semi-invariants and their relation to moments in which I will state four important known properties of theirs which are fundamental in the developments that follow. I will also develop the necessary and hitherto lacking explicit relation connecting semi-invariants and moments.

Given a statistical series, $x_1, x_2, x_3, \dots, x_N$, consisting of N recorded values of the same variate x , it has been customary to use certain symmetric functions of these N values, called moments, for the purpose of characterizing the series or distribution. The n th moment about the point from which x is measured is defined :

$$\nu'_n = \sum_{i=1}^N x_i^n / N. \quad (1)$$

If x varies continuously, then for the totality of possible values of x , the n th moment about the origin for x is defined:

$$\mu'_n = \int_{-\infty}^{\infty} x^n f(x) dx / \int_{-\infty}^{\infty} f(x) dx \quad (2)$$

where $f(x)$ is the probability function for x , i. e.,

$$f(x) dx / N$$

is the probability that a value of x chosen at random will fall between the limits

$$x \pm \frac{dx}{2}$$

With this definition it is to be observed that the integral in the denominator of (2) is equal to N .

The definition (2) is the one usually given for moments but in the case of the probability function it is better to choose the unit of measure for $f(x)$ in such a way that

$$f(x) dx$$

itself gives the probability that a value of x chosen at random will fall between the limits

$$x \pm \frac{dx}{2}.$$

In this case, evidently,

$$\int_{-\infty}^{\infty} f(x) dx = 1,$$

and

$$\mu'_n = \int_{-\infty}^{\infty} x^n f(x) dx \quad (2')$$

Moments defined in this way are said to be unifrequency moments and are the moments used in all that follows.

Thiele ⁽¹⁾ suggested another set of symmetric functions of the N values of x to be used instead of moments for the characterization of frequency functions which possess properties which make them of great advantage in many problems. These functions he called semi-invariants. For a set of N values of a discrete variate, x , they are defined:

$$N e^{\lambda_1 t + \frac{1}{2!} \lambda_2 t^2 + \dots} = \sum_{i=1}^N e^{x_i t} \quad (3)$$

in which λ_n is the n th semi-invariant of the frequency function for x . For the totality of possible values of a continuous variate, x , they are given by

$$e^{\lambda_1 t + \frac{1}{2!} \lambda_2 t^2 + \frac{1}{3!} \lambda_3 t^3 + \dots} = \int_{-\infty}^{\infty} dx f(x) e^{xt}. \quad (4)$$

(1) THIELE, T. N., Loc. cit.

Equations (3) and (4) are both identities in the arbitrary parameter t . It is obvious that the semi-invariants are functions of the moments and vice-versa. Equations connecting moments and semi-invariants are obtained by equating the coefficients of equal powers of t in (3) and (4).

For greater familiarity with the method of semi-invariants and their properties the reader is advised to consult the works of THIELE⁽¹⁾, CHARLIER⁽²⁾, WICKSELL⁽³⁾; and JORGENSEN⁽⁴⁾. In particular I will only mention here the four properties which are perhaps the most useful and remarkable. The first is the property of all the semi-invariants of order higher than the first of being independent of the origin from which the values of x are measured. The second is that if a variate x is itself the sum of any number of independent variates, each with its own probability function, then the semi-invariants of x are respectively the sums of the corresponding semi-invariants of the component variates⁽⁵⁾. The third is the remarkable reciprocal or inversion theorem, which is: If $f(x)$ is a frequency function (probability function), and if

$$e^{g(wi)} = \int_{-\infty}^{\infty} dx f(x) e^{xwi} = U(w) \quad (i = \sqrt{-1})$$

then

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dw e^{g(wi)} e^{-xwi} \quad (5)$$

and

$$e^{g(wi)} = e^{\lambda_1 wi} + \frac{1}{2!} \lambda_2 (wi)^2 + \frac{1}{3!} \lambda_3 (wi)^3 + \dots$$

⁽¹⁾ THIELE, T. N. Loc. cit., and other works. See bibliography at end of paper.

⁽²⁾ CHARLIER, C. V. L., Various papers published in the « Arkiv for Matematik, Astronomi och Fysik », Stockholm. In particular see *Frequency Curves of Type A in Heterogeneous Statistics*, Band 9, No. 25 (1914).

⁽³⁾ WICKSELL, S. D., Various papers published in the Arkiv mentioned above and also in the « Kungl. Svenska Vetenskapsakademiens Handlingar ». In particular see, *The Correlation Function of Type A*, in the latter publication, Band 58, No. 3 (1917).

⁽⁴⁾ JORGENSEN, N. R., *Undersøgelser over Frekvensflader og Korrelation*; Arnold Busck, Copenhagen (1916).

⁽⁵⁾ THIELE, T. N., Loc. cit., p. 39.

which means that if the semi-invariants of x are known then (5) gives $f(x)$. WICKSELL calls $U(w)$ the reciprocal function of $f(x)$. If the right-hand side of (5) is rewritten

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dw e^{-(x-\lambda_1)wi + \frac{\lambda_2(wi)^2}{2}} \left[1 - A_3(wi)^3 + A_4(wi)^4 + \dots \right] \quad (6)$$

the A 's are the coefficients in CHARLIER's Type A series⁽¹⁾ and these up to the fifth coincide with the semi-invariants of x . It is obvious that the conditions for convergence on $f(x)$ for (5) and (6) are not the same in terms of the λ 's as they are in terms of the A 's. The no satisfactory answer to the questions of convergence raised by (5) or (6) has been given from the theoretical point of view, it is to be expected, I think, that for this investigation the λ 's will be found more useful than either the moments or the A 's.

The fourth of these properties which I shall use is an extension of the reciprocal theorem. It may be stated as follows: Let $X_1, X_2, \dots, X_s$ be s variates with the correlation function $F(X_1, X_2, \dots, X_s)$, then the frequency function $P(z)$ of any given function $h(X_1, X_2, \dots, X_s)$ of the same variates is given by

$$P(z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dw e^{-zw} U(w)$$

where

$$U(w) = \int \int \dots \int_{-\infty}^{\infty} dX_1 dX_2 \dots dX_s F(X_1, X_2, \dots, X_s) e^{h(X_1, X_2, \dots, X_s)wi}$$

If the s variates are independent, each with its own probability function, the theorem still holds with the change that $F(X_1, X_2, \dots, X_s)$ is now replaced by the product of the s separate frequency functions⁽²⁾.

The definition of semi-invariants is readily extended for frequency functions of two or more variates. For instance, if $f(x, y)$ is the probability function of the variates, x and y , that is if

$$f(x, y) dx dy.$$

(1) CHARLIER, C. V. L., Loc. cit.

(2) CHARLIER, C. V. L., *Contributions to the Mathematical Theory of Statistics* 3, 4; « Arkiv for Matematik, Astronomi och Fysik » (Stockholm), Band 8, No. 4 (1912).

expresses the probability that a value of x and a value of y chosen at random simultaneously fall within the limits

$$x \pm \frac{1}{2} dx, \quad y \pm \frac{1}{2} dy,$$

then the semi-invariants of x and y are defined by

$$\begin{aligned} e^{(\lambda_{10}s + \lambda_{01}t)} + \frac{1}{2!}(\lambda_{10}s + \lambda_{01}t)^{(2)} + \frac{1}{3!}(\lambda_{10}s + \lambda_{01}t)^{(3)} + \dots \\ = \int \int_{-\infty}^{\infty} dx dy f(x, y) e^{(xs + yt)} \end{aligned} \quad (7)$$

for continuous variates, in which

$$(\lambda_{10}s + \lambda_{01}t)^{(n)}$$

denotes a symbolic modification of the binomial expansion, the meaning of which is made clear by the following illustration:

$$(\lambda_{10}s + \lambda_{01}t)^{(3)} = \lambda_{30}s^3 + 3\lambda_{21}s^2t + 3\lambda_{12}st^2 + \lambda_{03}t^3.$$

Needless to say, λ_{rs} is one of the semi-invariants of the $(r+s)$ th order of the distribution function for x and y , or as is commonly said, for the frequency surface whose equation is

$$z = f(x, y).$$

Again, if in (7) s and t be replaced respectively by si and ti , then the left-hand side may set equal to a $U(s, t)$ which will then be the reciprocal function of $f(x, y)$ and then by the extension of the inversion theorem to the case of two variates,

$$f(x, y) = \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} ds dt U(s, t) e^{-(xsi + yti)} \quad (8)$$

If the right-hand side be rewritten as follows,

$$\begin{aligned} f(x, y) = \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} ds dt e^{-[x-\lambda_{10}]si + [y-\lambda_{01}]ti} + \frac{1}{2}(\lambda_{10}si + \lambda_{01}ti)^{(2)} \\ \times [1 - \{A_{30}(si)^3 + A_{21}s^2ti^3 + A_{12}st^2i^3 + A_{03}(ti)^3\} + \dots] \end{aligned} \quad (9)$$

the A 's are the coefficients in the CHARLIER Type A series for two variates. It is clear from the above example for two variates how to

and that

$$\mu'_n = (-1)^n \begin{vmatrix} -\binom{0}{0} \lambda_1 & 1 & 0 & 0 & \dots & 0 \\ -\binom{1}{0} \lambda_2 & -\binom{1}{1} \lambda_1 & 1 & 0 & \dots & 0 \\ -\binom{2}{0} \lambda_3 & -\binom{2}{1} \lambda_2 & -\binom{2}{2} \lambda_1 & 1 & \dots & 0 \\ -\binom{3}{0} \lambda_4 & -\binom{3}{1} \lambda_3 & -\binom{3}{2} \lambda_2 & -\binom{3}{3} \lambda_1 & \dots & 0 \\ -\binom{4}{0} \lambda_5 & -\binom{4}{1} \lambda_4 & -\binom{4}{2} \lambda_3 & -\binom{4}{3} \lambda_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -\binom{n-1}{0} \lambda_n & -\binom{n-1}{1} \lambda_{n-1} & -\binom{n-1}{2} \lambda_{n-2} & -\binom{n-1}{3} \lambda_{n-3} \dots & -\binom{n-1}{n-1} \lambda_1 \end{vmatrix}$$

The expansions of these two determinants are accomplished in the same way; on account of a slight advantage in simplicity I shall deal with the latter first. Examination of the results for low values of n suggested that,

$$\begin{aligned} \mu'_n &= \lambda_n + \binom{n}{1} \lambda_{n-1} \lambda_1 + \binom{n}{2} \lambda_{n-2} \lambda_2 + \dots + \binom{n}{r} \lambda_{n-r} \lambda_r + \dots \\ &+ \frac{n!}{(n-4)! (2!)^2 2!} \lambda_{n-4} \lambda_2^2 + \dots + \frac{n!}{(a!)^r (b!)^s (c!)^t \dots r! s! t! \dots} \lambda_a^r \lambda_b^s \lambda_c^t \dots \\ &\quad (ar + bs + ct + \dots = n) \end{aligned}$$

with no combination of values of $a, b, c, \dots$ repeated, or that

$$\mu'_n = \Sigma \Sigma \Sigma \dots \frac{n!}{(a!)^r (b!)^s (c!)^t \dots r! s! t! \dots} \lambda_a^r \lambda_b^s \lambda_c^t \dots \quad (10)$$

where

$$a > b > c > \dots$$

and

$$ar + bs + ct + \dots = n$$

To complete the proof by induction, expand by minors with regard to the elements of the last row, assuming that the law holds for $\mu'_{n-1}, \mu'_{n-2}, \dots, \mu'_1$.

It is seen that the minor of $-\binom{n-1}{a} \lambda_{n-a}$ is simply μ'_a and that all signs come positive. Then consider the composition of the term in

$$\lambda_a^r \lambda_b^s \lambda_c^t \dots$$

with

$$a > b > c > \dots$$

$$ar + bs + ct + \dots = n$$

in the final result; it is evidently obtained as the sum of the products of $\binom{n-1}{n-a} \lambda_a$ by the term in its minor, i. e., in μ'_{n-a} , in $\lambda_a^{r-1} \lambda_b^s \lambda_c^t \dots$, of $\binom{n-1}{n-b} \lambda_b$ by the term in its minor in $\lambda_a^r \lambda_b^{s-1} \lambda_c^t \dots$, of $\binom{n-1}{n-c} \lambda_c$ by the term in its minor in $\lambda_a^r \lambda_b^s \lambda_c^{t-1} \dots$, etc. Whence the sought coefficient is then,

$$\begin{aligned} & \binom{n-1}{n-a} \frac{(n-a)!}{(a!)^{r-1} (b!)^s (c!)^t \dots (r-1)! s! t! \dots} + \\ & + \binom{n-1}{n-b} \frac{(n-b)!}{(a!)^r (b!)^{s-1} (c!)^t \dots r! (s-1)! t! \dots} + \dots \\ & = \frac{(n-1)!}{(a!)^r (b!)^s (c!)^t \dots r! s! t! \dots} \left[\frac{r a!}{(a-1)!} + \frac{s b!}{(b-1)!} + \frac{t c!}{(c-1)!} + \dots \right] \\ & = \frac{n!}{(a!)^r (b!)^s (c!)^t \dots r! s! t! \dots} \end{aligned}$$

since the term in the brackets reduces to $ar + bs + ct + \dots = n$. Thus the law holds for any value of n if it holds for all lower values.

The law for the expansion of λ_n in terms of μ' 's is

$$\begin{aligned} \lambda_n &= \mu'_n - \frac{n!}{(n-1)! 1!} \mu'_{n-1} \mu'_1 - \frac{n!}{(n-2)! 2!} \mu'_{n-2} \mu'_2 + \\ & + \frac{2! n!}{(n-2)! 2!} \mu'_{n-2} \mu'_1{}^2 - \frac{3! n!}{(n-3)! 3!} \mu'_{n-3} \mu'_1{}^3 + \dots \\ & + \frac{(-1)^{(r+s+t+\dots)-1} [(r+s+t+\dots)-1]! n!}{(a!)^r (b!)^s (c!)^t \dots r! s! t! \dots} \mu'_a{}^r \mu'_b{}^s \mu'_c{}^t \dots + \dots \end{aligned}$$

in which, as before,

$$a > b > c > \dots$$

and

$$ar + bs + ct + \dots = n.$$

This can also be compactly expressed:

$$\lambda_n = \Sigma \Sigma \Sigma \dots \frac{(-1)^{(r+s+t+\dots)-1} [(r+s+t+\dots)-1]! n! \mu'_a{}^r \mu'_b{}^s \mu'_c{}^t \dots}{(a!)^r (b!)^s (c!)^t \dots r! s! t! \dots} \quad (11)$$

with the same restrictions on the $a, b, c, \dots$ and $r, s, t, \dots$ as before. The proof proceeds just as in the first case. The law is assumed to hold for $\lambda_{n-1}, \lambda_{n-2}, \dots, \lambda_1$. Then on expansion by minors with reference to the elements of the last row, it is seen that every minor is precisely a λ of lower order and that the product of every element of the last row, except μ'_n , by its minor is negative. The coefficient of μ'_n in the expansion of the determinant is obviously unity which is also given by the formula to be established. And the coefficient of any other term $\mu'_a{}^r \mu'_b{}^s \mu'_c{}^t \dots$ in the expansion of the determinant is

$$\begin{aligned} & - \binom{n-1}{n-1-a} \frac{(n-a)! (-1)^{(r+s+t+\dots)-2} [(r+s+t+\dots)-2]!}{(a!)^{r-1} (b!)^s (c!)^t \dots (r-1)! s! t! \dots} \\ & - \binom{n-1}{n-1-b} \frac{(n-b)! (-1)^{(r+s+t+\dots)-2} [(r+s+t+\dots)-2]!}{(a!)^r (b!)^{s-1} (c!)^t \dots r! (s-1)! t! \dots} - \dots \\ & = \frac{(n-1)! (-1)^{(r+s+t+\dots)-1} [(r+s+t+\dots)-2]!}{(a!)^r (b!)^s (c!)^t \dots r! s! t! \dots} \times \\ & \quad \left[r(n-a) + s(n-b) + t(n-c) + \dots \right] \end{aligned}$$

The term in the brackets is

$$\begin{aligned} & rn - ra + sn - sb + tn - tc + \dots \\ & = n(r+s+t+\dots) - (ra + sb + tc + \dots) \\ & = n[(r+s+t+\dots) - 1] \end{aligned}$$

which is the result required.

By way of illustration of these two formulas and for future reference in this paper I list below some results obtained from them.

$$\begin{aligned} \lambda_1 &= \mu'_1 \\ \lambda_2 &= \mu'_2 - \mu'_1{}^2 \\ \lambda_3 &= \mu'_3 - 3\mu'_2 \mu'_1 + 2\mu'_1{}^3 \\ \lambda_4 &= \mu'_4 - 4\mu'_3 \mu'_1 - 3\mu'_2{}^2 + 12\mu'_2 \mu'_1{}^2 - 6\mu'_1{}^4 \\ \lambda_5 &= \mu'_5 - 5\mu'_4 \mu'_1 - 10\mu'_3 \mu'_2 + 20\mu'_3 \mu'_1{}^2 + 30\mu'_2{}^2 \mu'_1 \\ & \quad - 60\mu'_2 \mu'_1{}^3 + 24\mu'_1{}^5 \\ \lambda_6 &= \mu'_6 - 6\mu'_5 \mu'_1 - 15\mu'_4 \mu'_2 + 30\mu'_4 \mu'_1{}^2 - 10\mu'_3{}^2 + 120\mu'_3 \mu'_2 \mu'_1 \\ & \quad - 120\mu'_3 \mu'_1{}^3 + 30\mu'_2{}^3 - 180\mu'_2{}^2 \mu'_1{}^2 + 360\mu'_2 \mu'_1{}^4 - 120\mu'_1{}^6. \end{aligned} \quad (12)$$

In case the origin from which values of x are measured is chosen at the mean of those values, $\lambda_1 = \mu'_1 = 0$, and then :

$$\mu_1 = \lambda_1 = 0$$

$$\mu_2 = \lambda_2$$

$$\mu_3 = \lambda_3$$

$$\mu_4 = \lambda_4 + 3\lambda_2^2 \quad (13)$$

$$\mu_5 = \lambda_5 + 10\lambda_3\lambda_2$$

$$\mu_6 = \lambda_6 + 15\lambda_4\lambda_2 + 10\lambda_3^2 + 15\lambda_3^2$$

$$\mu_7 = \lambda_7 + 21\lambda_5\lambda_2 + 35\lambda_4\lambda_3 + 105\lambda_3\lambda_2^2$$

$$\mu_8 = \lambda_8 + 28\lambda_6\lambda_2 + 56\lambda_5\lambda_3 + 35\lambda_4^2 + 210\lambda_4\lambda_2^2 + 280\lambda_3^2\lambda_2 + 105\lambda_4^2$$

$$\mu_9 = \lambda_9 + 36\lambda_7\lambda_2 + 84\lambda_6\lambda_3 + 126\lambda_5\lambda_4 + 378\lambda_5\lambda_2^2 + 1260\lambda_4\lambda_3\lambda_2 + 280\lambda_3^3 + 1260\lambda_3\lambda_3^2$$

$$\mu_{10} = \lambda_{10} + 45\lambda_8\lambda_2 + 120\lambda_7\lambda_3 + 210\lambda_6\lambda_4 + 126\lambda_5^2 + 630\lambda_6\lambda_2^2 + 2520\lambda_5\lambda_3\lambda_2 + 1575\lambda_4^2\lambda_2 + 2100\lambda_4\lambda_3^2 + 3150\lambda_4\lambda_3^2 + 6300\lambda_3^2\lambda_2^2 + 945\lambda_5^2$$

$$\mu_{11} = \lambda_{11} + 55\lambda_9\lambda_2 + 165\lambda_8\lambda_3 + 330\lambda_7\lambda_4 + 462\lambda_6\lambda_5 + 990\lambda_7\lambda_2^2 + 4620\lambda_6\lambda_3\lambda_2 + 6930\lambda_5\lambda_4\lambda_2 + 4620\lambda_5\lambda_3^2 + 5775\lambda_4^2\lambda_3 + 6930\lambda_5\lambda_3^2 + 34650\lambda_4\lambda_3\lambda_2^2 + 15400\lambda_3^3\lambda_2 + 17325\lambda_3\lambda_4^2$$

$$\mu_{12} = \lambda_{12} + 66\lambda_{10}\lambda_2 + 220\lambda_9\lambda_3 + 495\lambda_8\lambda_4 + 792\lambda_7\lambda_5 + 462\lambda_6^2 + 1485\lambda_8\lambda_2^2 + 7920\lambda_7\lambda_3\lambda_2 + 13860\lambda_6\lambda_4\lambda_2 + 9240\lambda_6\lambda_3^2 + 8316\lambda_5^2\lambda_2 + 27720\lambda_5\lambda_4\lambda_3 + 5775\lambda_4^3 + 13860\lambda_6\lambda_3^2 + 83160\lambda_5\lambda_3\lambda_2^2 + 51975\lambda_4^2\lambda_2^2 + 138600\lambda_4\lambda_3^2\lambda_2 + 15400\lambda_4^4 + 51975\lambda_4\lambda_4^2 + 138600\lambda_3^2\lambda_3^2 + 10395\lambda_5^6$$

CHAPTER II

Semi-invariants of Moments about a Fixed Point

In order to give the reader greater familiarity with the method of semi-invariants applied to problems of sampling and to show its power in that direction I shall show how it is applied to a problem that has been successfully dealt with in other ways. The elegant solution given below is due to Professor S. D. WICKSELL and is taken from his unpublished notes with his permission.

The formulation of the problem is as follows: From an infinite parent distribution a value of an x_1 , a value of an x_2 , a value of an x_3 , ..., and a value of an x_N are each chosen at random and independently of each other. The process is repeated indefinitely many times. We wish to find the semi-invariants of the distribution function of values of the power-sum $(x_1^r + x_2^r + \dots + x_N^r)/N$, which is, of

course, the r -th unifrequency moment of a sample of N or a sample r -th moment, given the moments, $\mu'_1, \mu'_2, \mu'_3, \dots$ of the parent distribution.

Then, by definition, the required semi-invariants, $b_1, b_2, b_3, \dots$ are given by the equation,

$$e^{b_1 t + \frac{1}{2!} b_2 t^2 + \frac{1}{3!} b_3 t^3 + \dots} = \int_{-\infty}^{\infty} dx_1 \dots \int_{-\infty}^{\infty} dx_N f_1(x_1) \dots f_N(x_N) e^{\left[\frac{x_1^r + x_2^r + \dots + x_N^r}{N} - \mu'_r \right] t}$$

in which $f_s(x_s)$ is the distribution function of values of x_s . The right-hand side can be rewritten,

$$\int_{-\infty}^{\infty} dx_1 f_1(x_1) e^{\left(\frac{x_1^r}{N} - \frac{\mu'_r}{N} \right) t} \cdot \int_{-\infty}^{\infty} dx_2 f_2(x_2) e^{\left(\frac{x_2^r}{N} - \frac{\mu'_r}{N} \right) t} \dots \int_{-\infty}^{\infty} dx_N f_N(x_N) e^{\left(\frac{x_N^r}{N} - \frac{\mu'_r}{N} \right) t}$$

Now if $f(x)$ be the distribution function for the parent distribution, then obviously,

$$f_1(x_1) = f_2(x_2) = \dots = f_N(x_N),$$

and the above can be replaced by,

$$e^{b'_1 t + \frac{1}{2!} b'_2 t^2 + \frac{1}{3!} b'_3 t^3 + \dots} e^{b'_1 t + \frac{1}{2!} b'_2 t^2 + \frac{1}{3!} b'_3 t^3 + \dots} \dots e^{b'_1 t + \frac{1}{2!} b'_2 t^2 + \frac{1}{3!} b'_3 t^3 + \dots}$$

in which the definition of b'_s is apparent. Then finally,

$$e^{b_1 t + \frac{1}{2!} b_2 t^2 + \frac{1}{3!} b_3 t^3 + \dots} = e^{N(b'_1 t + \frac{1}{2!} b'_2 t^2 + \frac{1}{3!} b'_3 t^3 + \dots)}$$

The values of the b'_s 's and hence of the b_s 's are readily written out by means of the known law connecting moments and semi-invariants. Thus from the first five of the relations (12) there follow at once, on observing that $b_1 = \mu'_r - \mu'_r = 0$,

$$b_1 = 0$$

$$b_2 = \frac{1}{N} \int_{-\infty}^{\infty} dx f(x) (x^r - \mu'_r)^2 = \frac{1}{N} (\mu'_{2r} - \mu'^r_r).$$

$$\begin{aligned}
b_3 &= \frac{1}{N^2} \left[\mu'_{3r} - 3 \mu'_{2r} \mu'_r + 2 \mu'^3_r \right] \\
b_4 &= \frac{1}{N^3} \left[(\mu'_{4r} - 4 \mu'_{3r} \mu'_r + 6 \mu'_{2r} \mu'^2_r - 3 \mu'^4_r) - 3 (\mu'_{2r} - \mu'^2_r)^2 \right] \\
&= \frac{1}{N^3} \left(\mu'_{4r} - 4 \mu'_{3r} \mu'_r - 3 \mu'^2_{2r} + 12 \mu'_{2r} \mu'^2_r - 6 \mu'^4_r \right) \quad (14) \\
b_5 &= \frac{1}{N^4} \left(\mu'_{5r} - 5 \mu'_{4r} \mu'_r - 10 \mu'_{3r} \mu'_{2r} + 20 \mu'_{3r} \mu'^2_r + 30 \mu'^2_{2r} \mu'_r \right. \\
&\quad \left. - 60 \mu'_{2r} \mu'^3_r + 24 \mu'^5_r \right)
\end{aligned}$$

Here the moments are computed about a *fixed point*, the origin from which the x 's of the parent distribution are counted.

An easy generalization is effected by supposing the x_1 , the $x_2, \dots$, and the x_N , each to be drawn from separate and distinct parent distributions, each with its own frequency function.

An interesting particular case is that in which $r = 2$ and the parent distribution is normal. Here it can be readily shown that

$$b_k = \left(\frac{2}{N} \right)^{k-1} b'^{k-1}_{2} (k-1)! (b'_2 - k b'^2_1)$$

In particular if $b'_1 = 0$,

$$b_k = \left(\frac{2}{N} \right)^{k-1} b'^k_2 (k-1)!$$

that is, the error function for v'_2 taken around the mean of the parent distribution is a Pearson's Type III curve! We get in this case

$$F(v'_2) = \frac{(v'_2)^{N/2-1} e^{-\frac{v'_2}{b'_2}}}{(2b'_2)^{N/2} \Gamma(N/2)}$$

as has already been shown by HELMERT ⁽¹⁾. A proof based on the present method is easily carried thru by induction in the same way that the formula for λ_n in terms of moments is found, that is, by noting how λ_n is made up of products of binomial coefficients into λ 's of lower orders.

(1) HELMERT, F. R., « Zeitschrift für Mathematik und Physik », Vol. XXI, (1876) p. 202.

See also CZUBER, E., *Theorie der Beobachtungsfehler*, Leipzig (1891) p. 149.

CHAPTER III

The Semi-invariants of Moments about the Mean

The problem to be attacked is as follows: Let x be a variate with the probability function $f(x)$: N repetitions of x have been observed with the values, $x_1, x_2, \dots, x_N$, and the mean,

$$\sum_{i=1}^N \frac{x_i}{N} = v'_1,$$

has been computed. Setting

$$\delta_i = x_i - v'_1,$$

the values of the moments about the mean

$$v_n = \sum_{i=1}^N \frac{\delta_i^n}{N},$$

have next been computed. Required to find the semi-invariants of the frequency function for v_n , for particular values of n , for infinitely many sets of N repetitions of x , on the assumption that the semi-invariants of $f(x)$ are known, and that each of the N repetitions of x in each single set is independent of all the others.

A. *The Correlation Function of the δ 's, i. e., of Deviations from the Mean.*

The N δ 's in each set satisfy the relation,

$$\sum_{i=1}^N \delta_i = 0.$$

But let

$$F(\delta_1, \delta_2, \dots, \delta_{N-1})$$

be the correlation or probability function of the first $N - 1$ δ 's of each set which form a set of correlated variates in the usual sense of the word. That is, let

$$F(\delta_1, \delta_2, \dots, \delta_{N-1}) d\delta_1 d\delta_2 \dots d\delta_{N-1}$$

give the probability that the first $N - 1$ δ 's simultaneously fall within the limits,

$$\delta_1 \pm \frac{1}{2} d\delta_1, \delta_2 \pm \frac{1}{2} d\delta_2, \dots, \delta_{N-1} \pm \frac{1}{2} d\delta_{N-1}$$

respectively.

Then the semi-invariants of $F(\delta_1, \delta_2, \dots, \delta_{N-1})$ are defined by,

$$e^{\left(\sum_{i=1}^{N-1} \lambda_i t_i\right) + \frac{1}{2!} \left(\sum_{i=1}^{N-1} \lambda_i t_i\right)^{(2)} + \frac{1}{3!} \left(\sum_{i=1}^{N-1} \lambda_i t_i\right)^{(3)} + \dots} = \int_{-\infty}^{\infty} d\delta_1 \int_{-\infty}^{\infty} d\delta_2 \dots \int_{-\infty}^{\infty} d\delta_{N-1} F(\delta_1, \delta_2, \dots, \delta_{N-1}) e^{\left(\sum_{i=1}^{N-1} \delta_i t_i\right)} \quad (15)$$

in which

$$\left(\sum_{i=1}^{N-1} \lambda_i t_i\right)^{(r)}$$

denotes a symbolic multinomial expansion; its meaning is clear from the example,

$$\left(\sum_{i=1}^2 \lambda_i t_i\right)^{(2)} = \lambda_{20} t_1^2 + 2 \lambda_{11} t_1 t_2 + \lambda_{02} t_2^2,$$

and in which

$$e^{\left(\sum_{i=1}^{N-1} \delta_i t_i\right)}$$

is expanded in an analogous manner. In fact, the right-hand side becomes after expansion,

$$1 + \left(\sum_{i=1}^{N-1} v_i t_i\right) + \frac{1}{2!} \left(\sum_{i=1}^{N-1} v_i t_i\right)^{(2)} + \frac{1}{3!} \left(\sum_{i=1}^{N-1} v_i t_i\right)^{(3)} + \dots$$

and values of λ 's in terms of v 's are found by equating coefficients of like powers of the t 's. It is to be noted that

$$v_{10\dots 0} = v_{010\dots 0} = \dots = v_{00\dots 01} = \lambda_{10\dots 0} = \lambda_{010\dots 0} = \dots = \lambda_{00\dots 01}.$$

Now set

$$\delta_i = \sum_{j=1}^N a_{ij} x_j, \quad \text{with} \quad \begin{cases} a_{ij} = -1/N, & i \neq j \\ a_{ii} = \frac{N-1}{N} \end{cases}$$

and it follows at once that,

$$\begin{aligned} & e^{\left(\sum_{i=1}^{N-1} \lambda_i t_i\right) + \frac{1}{2!} \left(\sum_{i=1}^{N-1} \lambda_i t_i\right)^{(2)} + \frac{1}{3!} \left(\sum_{i=1}^{N-1} \lambda_i t_i\right)^{(3)} + \dots} \\ &= \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2 \dots \int_{-\infty}^{\infty} dx_N f(x_1) \cdot f(x_2) \dots f(x_N) e^{\sum_{i=1}^{N-1} \sum_{j=1}^N a_{ij} x_j t_i} \quad (16) \\ &= e^{\lambda_1 \left(\sum_{i=1}^{N-1} a_{i1} t_i\right) + \frac{\lambda_2}{2!} \left(\sum_{i=1}^{N-1} a_{i1} t_i\right)^{(2)} + \dots} \\ & \quad e^{\lambda_1 \left(\sum_{i=1}^{N-1} a_{i2} t_i\right) + \frac{\lambda_2}{2!} \left(\sum_{i=1}^{N-1} a_{i2} t_i\right)^{(2)} + \dots} \end{aligned}$$

in which $\lambda_1, \lambda_2, \dots$ are the semi-invariants of $f(x)$. Thus, then,

$$\left(\sum_{i=1}^{N-1} \lambda_i t_i \right)^{(k)} = \lambda_k \left[\left(\sum_{i=1}^{N-1} a_{i1} t_i \right)^k + \left(\sum_{i=1}^{N-1} a_{i2} t_i \right)^k + \dots + \left(\sum_{i=1}^{N-1} a_{iN} t_i \right)^k \right] \quad (16a)$$

or

$$\frac{k!}{k_1! k_2! \dots k_{N-1}!} \lambda_{k_1 k_2 \dots k_{N-1}} = \lambda_k \frac{k!}{k_1! k_2! \dots k_{N-1}!} \times \left[a_{11}^{k_1} a_{21}^{k_2} \dots a_{N-1,1}^{k_{N-1}} + a_{21}^{k_1} a_{22}^{k_2} \dots a_{N-1,2}^{k_{N-1}} + \dots \right]$$

where

$$k_1 + k_2 + \dots + k_{N-1} = k,$$

which gives

$$\begin{aligned} j_{k_1 k_2 k_3 \dots k_{N-1}} &= \lambda_k \sum_{j=1}^N a_{j1}^{k_1} a_{j2}^{k_2} \dots a_{j,N-1}^{k_{N-1}} \\ &= \lambda_k \sum_{j=1}^N a_{1j}^{k_1} a_{2j}^{k_2} \dots a_{N-1,j}^{k_{N-1}} \quad (17) \\ &= \lambda_k \frac{1}{N^k} \left[\sum_{i=1}^{N-1} (-1)^{k-i} (N-1)^{k_i} + \dots (-1)^k \right] \end{aligned}$$

The summation, of course, refers to the various k 's which appear in the subscript on the left-hand side. The essential trick in this development is the expression of the δ 's as linear functions of the N independent x 's ⁽¹⁾

On account of the relation,

$$\sum_{i=1}^N \delta_i = 0,$$

the correlation function,

$$F(\delta_1, \delta_2, \dots, \delta_{N-1}, \delta_N)$$

does not exist in the usual sense of the word. But there is a probability that a certain set of N δ 's will occur in a given order and not another set. This probability is the probability that the first $N-1$ of the δ 's occur for, then the appearance of the last is a certainty. But it is found that if one proceeds as if there were a correlation function,

$$F(\delta_1, \delta_2, \dots, \delta_{N-1}, \delta_N)$$

and uses it in (15) and in (16) in the same way that $F(\delta_1, \dots, \delta_{N-1})$ is used that one gets true results, that the marginal semi-invariants

⁽¹⁾ See THIELE, T. N. Loc. cit., p. 36.

of the two functions agree as they should. Why is this? The answer is that one does not stop with (15) but proceeds to (16) and it is valid in either case. But now for $F(\delta_1, \delta_2, \dots, \delta_{N-1}, \delta_N)$, keeping the above remarks in mind the formula (17) can be written in a more compact form:

$$\lambda_{k_1 k_2 \dots k_N} = \frac{\lambda_k}{N^k} \left[\sum_{i=1}^N (-1)^{k-k_i} (N-1)^{k_i} \right] \quad (17)$$

This gives the necessary relation between the semi-invariants of f and F .

All semi-invariants $\lambda_{k_1, k_2 \dots k_N}$ of F whose subscripts are permutations of the same set of N numbers shall be said to be of the same type. It follows at once that all semi-invariants of the same type are equal as may be seen from the above.

It may be remarked here that the method used above is also adapted to the case in which the set $x_1, x_2, \dots, x_N$, are observed values of N different variates each with its own distinct frequency function. However in order to reduce the results to a usable compass, relations between the N frequency functions would have to be postulated.

B. *The Frequency Function for $v_n = \sum_{i=1}^N \frac{\delta_i^n}{N}$.*

Let $P(v_n)$ be the probability function for v_n . Then the semi-invariants, S_r , of $P(v_n)$ are defined by,

$$\begin{aligned} e^{S_1 t + \frac{1}{2!} S_2 t^2 + \frac{1}{3!} S_3 t^3 + \dots} \\ = \int_{-\infty}^{\infty} d v_n P(v_n) e^{v_n t} \end{aligned} \quad (18)$$

The problem is now to express the semi-invariants, S_r , in terms of the semi-invariants, $\lambda_{lmn} \dots$, of the correlation function for δ 's found in the last section. For this purpose the right-hand side of (18) is replaced by,

$$\int_{-\infty}^{\infty} d \delta_1 \int_{-\infty}^{\infty} d \delta_2 \dots \int_{-\infty}^{\infty} d \delta_N F(\delta_1, \delta_2, \dots, \delta_N) e^{i \sum_{i=1}^N \delta_i^n t/N}$$

making use again of the fourth of the properties of semi-invariants mentioned in the first chapter. Next the semi-invariants $L_{rst} \dots$,

of the correlation function for n -th powers of the δ 's are defined by,

$$e^{\left(\sum_{i=1}^N L_i t_i\right)} + \frac{1}{2!} \left(\sum_{i=1}^N L_i t_i\right)^{(2)} + \frac{1}{3!} \left(\sum_{i=1}^N L_i t_i\right)^{(3)} + \dots$$

$$= \int_{-\infty}^{\infty} d\delta_1 \dots \int_{-\infty}^{\infty} d\delta_N F(\delta_1, \dots, \delta_N) e^{\left(\sum_{i=1}^N \delta_i^n t_i\right)} \quad (19)$$

Then the relation between the S 's and the L 's is immediately seen to be,

$$S_k = \frac{1}{N^k} \sum \frac{k!}{k_1! k_2! \dots k_r!} L_{k_1 k_2 \dots k_r}, \quad (20)$$

in which the summation is to be performed for all values of $k_1, k_2, \dots, k_r$, such that $k_1 + k_2 + \dots + k_r = k$.

Also we have on expansion of the right-hand side of (19),

$$e^{\left(\sum_{i=1}^N L_i t_i\right)} + \frac{1}{2!} \left(\sum_{i=1}^N L_i t_i\right)^{(2)} + \frac{1}{3!} \left(\sum_{i=1}^N L_i t_i\right)^{(3)} + \dots$$

$$= 1 + \left(\sum_{i=1}^N \nu_{ni} t_i\right) + \frac{1}{2!} \left(\sum_{i=1}^N \nu_{ni} t_i\right)^{(2)} + \frac{1}{3!} \left(\sum_{i=1}^N \nu_{ni} t_i\right)^{(3)} + \dots \quad (21)$$

This is the last link in the chain from the sought S'_s to the known λ 's of $f(x)$. But the means for expressing the S'_s directly and explicitly in terms of the semi-invariants of $f(x)$ are still lacking. The next sections deal with this shortcoming, showing the possibility of writing such an explicit relation and developing the practical method of calculating the S'_s in terms of the λ'_s of $f(x)$.

C. Explicit Relations between the S'_s and the Semi-invariants of the Correlation Function of the δ'_s .

The explicit relations (10) and (11) which give moments in terms of semi-invariants and vice-versa are immediately applicable in the calculation of the S'_s . Set in (21) and (15) they give,

$$\sum_{i=1}^N (L_i t_i)^{(n)} =$$

$$= \sum \sum \dots \frac{(-1)^{(r+s+t+\dots)-1} [(r+s+t+\dots)-1]! \left[\left(\sum_{i=1}^N \nu_{ni} t_i \right)^{(a)} \right]^r \left[\left(\sum_{i=1}^N \nu_{ni} t_i \right)^{(b)} \right]^s \dots}{(a!)^r (b!)^s (c!)^t \dots r! s! t! \dots} \quad (22)$$

and

$$\left(\sum_{i=1}^N \nu_i t_i \right)^{(n)} = \sum \sum \dots \frac{n! \left[\left(\sum_{i=1}^N \lambda_i t_i \right)^{(a)} \right]^r \left[\sum_{i=1}^N \lambda_i t_i \right]^{(b)} \dots}{(a!)^r (b!)^s (c!)^t \dots r! s! t!} \quad (23)$$

in which the L 's can be found in terms of the moments of F and these moments in turn found in terms of the semi-invariants of f by equating the coefficients of like powers of the t 's on both sides of the two equations. Examples of the applications of these two formulas follow:

$$\left(\sum_{i=1}^N L_i t_i\right)^{(3)} = \left(\sum_{i=1}^N v_{2i} t_i\right)^{(3)} - 3 \left(\sum_{i=1}^N v_{2i} t_i\right)^{(2)} \left(\sum_{i=1}^N v_{2i} t_i\right) + 2 \left(\sum_{i=1}^N v_{2i} t_i\right)^3 \quad (24)$$

gives

$$\begin{aligned} L_{30} \dots 0 &= v_{60} \dots 0 - 3 v_{40} \dots 0 v_{20} \dots 0 + 2 v_{20}^3 \dots 0 \\ L_{210} \dots 0 &= v_{420} \dots 0 - v_{40} \dots 0 v_{020} \dots 0 - 2 v_{220} \dots 0 v_{20} \dots 0 + 2 v_{20}^2 v_{020} \dots 0 \\ L_{1110} \dots 0 &= v_{2220} \dots 0 - v_{220} \dots 0 v_{0020} \dots 0 - v_{2020} \dots 0 v_{020} \dots 0 \\ &\quad - v_{0220} \dots 0 v_{20} \dots 0 + v_{20} \dots 0 v_{020} \dots 0 v_{0020} \dots 0 \end{aligned} \quad (24a)$$

and

$$\begin{aligned} \left(\sum_{i=1}^N v_i t_i\right)^{(6)} &= \left(\sum_{i=1}^N \lambda_i t_i\right)^{(6)} + 15 \left(\sum_{i=1}^N \lambda_i t_i\right)^{(4)} \left(\sum_{i=1}^N \lambda_i t_i\right)^{(2)} + \\ &\quad + 10 \left[\left(\sum_{i=1}^N \lambda_i t_i\right)^{(3)}\right]^2 + 15 \left[\left(\sum_{i=1}^N \lambda_i t_i\right)^{(2)}\right]^3 \end{aligned}$$

gives

$$\begin{aligned} v_{60} \dots 0 &= \lambda_{60} \dots 0 + 15 \lambda_{40} \dots 0 \lambda_{20} \dots 0 + 10 \lambda_{30}^2 \dots 0 + 15 \lambda_{20}^3 \dots 0 \\ v_{420} \dots 0 &= \lambda_{420} \dots 0 + \lambda_{40} \dots 0 \lambda_{020} \dots 0 + 8 \lambda_{310} \dots 0 \lambda_{110} \dots 0 + 6 \lambda_{220} \dots 0 \lambda_{20} \dots 0 \\ &\quad + 6 \lambda_{210}^2 \dots 0 + 4 \lambda_{30} \dots 0 \lambda_{120} \dots 0 + 3 \lambda_{20}^2 \dots 0 \lambda_{020} \dots 0 + 12 \lambda_{20} \dots 0 \lambda_{110}^2 \dots 0 \end{aligned}$$

In actual computation it is always kept in mind that all v 's and λ 's of the same type are equal and the work and the space needed are thereby greatly reduced.

However by substituting the expression for L 's given in (22) into the right-hand member of (20) a direct expression for S 's in terms of the moments of F is obtained. In fact, I get,

$$S_k(v_n) = \frac{1}{N^k} \Sigma \Sigma \dots \quad (25)$$

$$\dots \frac{(-1)^{(r+s+t+\dots)-1} [(r+s+t+\dots)-1]! k! v_{a_1 n, a_2 n, \dots}^r v_{b_1 n, b_2 n, \dots}^s v_{c_1 n, c_2 n, \dots}^t}{[a_1! a_2! \dots]^r [b_1! b_2! \dots]^s [c_1! c_2! \dots]^t \dots r! s! t! \dots}$$

the summation including all terms such that

$$\left\{ \begin{array}{l} r(a_1 + a_2 + \dots) + s(b_1 + b_2 + \dots) + t(c_1 + c_2 + \dots) + \dots = k \\ a_1 \geq a_2 \geq a_3 \geq \dots \\ b_1 \geq b_2 \geq b_3 \geq \dots \\ c_1 \geq c_2 \geq c_3 \geq \dots \\ \dots \dots \dots \dots \dots \dots \dots \\ (a_1 + a_2 + a_3 + \dots) > (b_1 + b_2 + \dots) > (c_1 + c_2 + \dots) > \dots \end{array} \right.$$

This follows since

$$S_k(v_n) = \frac{1}{N^k} \Sigma \dots \Sigma L_{k_1 k_2 k_3 \dots} \frac{k!}{k_1! k_2! k_3! \dots} = \frac{1}{N^k} \left(\sum_{i=1}^N L_i \right)^{(k)}$$

and

$$\begin{aligned} & \left(\sum_{i=1}^N L_i \right)^{(k)} = \\ & \frac{(-1)^{(r+s+t+\dots)-1} [(r+s+t+\dots)-1]! k! \left[\left(\sum_{i=1}^N v_{ni} \right)^{(a)} \right]^r \left[\left(\sum_{i=1}^N v_{ni} \right)^{(b)} \right]^s \left[\left(\sum_{i=1}^N v_{ni} \right)^{(c)} \right]^t \dots}{(a!)^r (b!)^s (c!)^t \dots r! s! t! \dots} \end{aligned} \quad (26)$$

merely by omitting the t 's which are not at all essential.

For future reference I give below the first five S 's for v_n obtained from (26). In these expressions and in the remainder of this paper zeros after the first in subscripts are omitted since no ambiguity is thereby introduced.

$$S_1(v_n) = v_{n,0} \quad (27)$$

$$S_2(v_n) = \frac{1}{N} \left[v_{2n,0} + (N-1) v_{n,n,0} - N v_{n,0}^2 \right]$$

$$S_3(v_n) = \frac{1}{N^2} \left[v_{3n,0} + 3(N-1) v_{2n,n,0} + (N-1)(N-2) v_{n,n,n,0} - 3N v_{2n,0} v_{n,0} - 3N(N-1) v_{n,n,0} v_{n,0} + 2N^2 v_{n,0}^3 \right]$$

$$\begin{aligned} S_4(v_n) = \frac{1}{N^3} & \left[v_{4n,0} + 4(N-1) v_{3n,n,0} + 3(N-1) v_{2n,2n,0} \right. \\ & + 6(N-1)(N-2) v_{2n,n,n,0} + (N-1)(N-2)(N-3) v_{n,n,n,n,0} \\ & - 4N v_{3n,0} v_{n,0} - 12N(N-1) v_{2n,n,0} v_{n,0} \\ & - 4N(N-1)(N-2) v_{n,n,n,0} v_{n,0} - 3N v_{2n,0}^2 + 12N^2 v_{2n,0} v_{n,0}^2 \\ & - 6N(N-1) v_{2n,0} v_{n,n,0} + 12N^2(N-1) v_{n,n,0} v_{n,0}^2 \\ & \left. - 3N(N-1)^2 v_{n,n,0}^2 - 6N^3 v_{n,0}^4 \right] \end{aligned}$$

$$\begin{aligned} S_5(v_n) = \frac{1}{N^4} & \left[v_{5n,0} + 5(N-1) v_{4n,n,0} + 10(N-1) v_{3n,2n,0} \right. \\ & + 10(N-1)(N-2) v_{3n,n,n,0} + 15(N-1)(N-2) v_{2n,2n,n,0} \\ & + 10(N-1)(N-2)(N-3) v_{2n,n,n,n,0} \\ & + (N-1)(N-2)(N-3)(N-4) v_{n,n,n,n,n,0} \\ & - 5N v_{4n,0} v_{n,0} - 20N(N-1) v_{3n,n,0} v_{n,0} \\ & - 15N(N-1) v_{2n,2n,0} v_{n,0} - 30N(N-1)(N-2) v_{2n,n,n,0} v_{n,0} \\ & \left. - 5N(N-1)(N-2)(N-3) v_{n,n,n,n,0} v_{n,0} - 10N v_{3n,0} v_{2n,0} \right] \end{aligned}$$

$$\begin{aligned}
& -10 N (N-1) v_{3n,0} v_{n,n,0} - 30 N (N-1) v_{2n,n,0} v_{2n,0} \quad (27) \\
& - 30 N (N-1)^2 v_{2n,n,0} v_{n,n,0} - 10 N (N-1) (N-2) v_{n,n,0} v_{2n,0} \\
& - 10 N (N-1)^2 (N-2) v_{n,n,0} v_{n,n,0} + 20 N^2 v_{3n,0} v_{n,0}^2 \\
& + 60 N^2 (N-1) v_{2n,n,0} v_{n,0}^2 + 20 N^2 (N-1) (N-2) v_{n,n,n,0} v_{n,0}^2 \\
& + 30 N^2 v_{2n,0}^2 v_{n,0} + 60 N^2 (N-1) v_{2n,0} v_{n,n,0} v_{n,0} \\
& + 30 N^2 (N-1)^2 v_{n,n,0}^2 v_{n,0} - 60 N^3 v_{2n,0} v_{n,0}^3 \\
& - 60 N^3 (N-1) v_{n,n,0} v_{n,0}^3 + 24 N^4 v_{n,0}^5 \Big].
\end{aligned}$$

In order to avoid numerical slips in these five results I have also obtained them by direct calculation as a check. The direct calculation grows somewhat tedious in the last.

In the expression (26) for $(\sum_{i=1}^N L_i)^{(k)}$ the $(\sum_{i=1}^N v_{ni})^{(p)}$'s could be replaced by their values in terms of $(\sum_{i=1}^N \lambda_i)^{(np)}$'s and these could again be explicitly expressed in terms of the semi-invariants of $f(x)$ by (16^a). Thus it is possible to write down the general formula for the m th semi-invariant of the n th moment but it is also possible to foresee that there is little to be gained by actually putting it together. It would be a super-compounding of (25) in which the law of formation of terms is already sufficiently complicated. Actual substitution in such a formula would have to be accomplished piecemeal; it would be equivalent in each case to a calculation of the kind that is now to be explained.

D. Actual Calculation of the S's.

As an example of the use of the formulas that have been developed for computing the S's, the calculation of $S_2(v_4)$ is here given in detail.

From (27),

$$S_2(v_4) = \frac{1}{N} [v_{80} \dots + (N-1) v_{440} \dots - N v_{40}^2 \dots].$$

and from the value of v_8 in terms of λ 's,

$$\begin{aligned}
(\sum v_i t_i)^{(8)} &= (\sum \lambda_i t_i)^{(8)} + 28 (\sum \lambda_i t_i)^{(6)} (\sum \lambda_i t_i)^{(2)} + 56 (\sum \lambda_i t_i)^{(5)} (\sum \lambda_i t_i)^{(3)} \\
&+ 35 [(\sum \lambda_i t_i)^{(4)}]^2 + 210 (\sum \lambda_i t_i)^{(4)} [(\sum \lambda_i t_i)^{(2)}]^2 \\
&+ 280 [(\sum \lambda_i t_i)^{(3)}]^2 (\sum \lambda_i t_i)^{(2)} + 105 [(\sum \lambda_i t_i)^{(2)}]^4.
\end{aligned}$$

(Summations are always with regard to i from 1 to N). In what follows only the essential parts of subscripts are retained. It is to be

remembered that all λ 's of the same type are equal. In the above, then, v_{80} is equal to the coefficient of t_1^8 on the right-hand side;

v_{44} to the coefficient of $t_1^4 t_2^4$ on the right-hand side divided by $\binom{8}{4}$.

This gives,

$$\begin{aligned} v_{80} &= \lambda_{80} + 28 \lambda_{60} \lambda_{20} + 56 \lambda_{50} \lambda_{30} + 35 \lambda_{40}^2 + 210 \lambda_{40} \lambda_{20}^2 + 280 \lambda_{30}^2 \lambda_{20} + 105 \lambda_{20}^4 \\ v_{44} &= \lambda_{44} + (12 \lambda_{42} \lambda_{20} + 16 \lambda_{33} \lambda_{11}) + (8 \lambda_{41} \lambda_{30} + 48 \lambda_{32} \lambda_{21}) + (\lambda_{40}^2 + 16 \lambda_{31}^2 \\ &\quad + 18 \lambda_{22}^2) + (6 \lambda_{40} \lambda_{20}^2 + 96 \lambda_{31} \lambda_{20} \lambda_{11} + 36 \lambda_{22} \lambda_{20}^2 + 72 \lambda_{22} \lambda_{11}^2) \\ &\quad + (16 \lambda_{30}^2 \lambda_{11} + 48 \lambda_{30} \lambda_{21} \lambda_{20} + 144 \lambda_{21}^2 \lambda_{11} + 72 \lambda_{21}^2 \lambda_{20}) + (9 \lambda_{20}^4 \\ &\quad + 72 \lambda_{20}^2 \lambda_{11}^2 + 24 \lambda_{11}^4). \end{aligned}$$

And in a similar way,

$$v_{40}^2 = \lambda_{40}^2 + 6 \lambda_{40} \lambda_{20}^2 + 9 \lambda_{20}^4.$$

Terms in v_{44} which will contribute to the same coefficients of semi-invariants of $f(x)$ in the final result are enclosed within parentheses. The final result will contain terms in λ_8 , λ_6 , λ_2 , $\lambda_5 \lambda_3$, λ_2^4 , $\lambda_3^2 \lambda_2$, and λ_2^4 together with their coefficients which are polynomials in N . Therefore in substituting the values of v_{80} , v_{44} , and v_{40}^2 in $S_2(v_4)$, the terms are arranged so that those contributing to the same coefficient in the final result are grouped together and enclosed within braces. Thus,

$$\begin{aligned} S_2(v_4) &= \frac{1}{N} \left[\left\{ \lambda_{80} + (N-1) \lambda_{44} \right\} + \left\{ 28 \lambda_{60} \lambda_{20} + (N-1) (12 \lambda_{42} \lambda_{20} \right. \right. \\ &\quad \left. \left. + 16 \lambda_{33} \lambda_{11}) \right\} + \left\{ 56 \lambda_{50} \lambda_{30} + (N-1) (8 \lambda_{41} \lambda_{30} + 48 \lambda_{32} \lambda_{21}) \right\} \\ &\quad + \left\{ 35 \lambda_{40}^2 + (N-1) (\lambda_{40}^2 + 16 \lambda_{31}^2 + 18 \lambda_{22}^2) - N \lambda_{20}^4 \right\} \\ &\quad + \left\{ 210 \lambda_{40} \lambda_{20}^2 + (N-1) (6 \lambda_{40} \lambda_{20}^2 + 96 \lambda_{31} \lambda_{20} \lambda_{11} + 36 \lambda_{22} \lambda_{20}^2 \right. \\ &\quad \left. + 72 \lambda_{22} \lambda_{11}^2) - 6 N \lambda_{40} \lambda_{20}^2 \right\} + \left\{ 280 \lambda_{30}^2 \lambda_{20} + (N-1) (16 \lambda_{30}^2 \lambda_{11} \right. \\ &\quad \left. + 48 \lambda_{30} \lambda_{21} \lambda_{20} + 144 \lambda_{21}^2 \lambda_{11} + 72 \lambda_{21}^2 \lambda_{20}) \right\} + \left\{ 105 \lambda_{20}^4 \right. \\ &\quad \left. + (N-1) (9 \lambda_{20}^4 + 72 \lambda_{20}^2 \lambda_{11}^2 + 24 \lambda_{11}^4) - 9 N \lambda_{20}^4 \right\} \Big]. \end{aligned}$$

Each term in (28) is now reduced in turn. The coefficient of λ_8 is

$$\begin{aligned} \frac{1}{N^8} &\left\{ [(N-1)^8 + N-1] + (N-1) [2(N-1)^4 + N-2] \right\} \\ &= \frac{(N-1)^2 (N^2 - 3N + 3)^2}{N^7} \end{aligned}$$

The coefficient of $\lambda_6 \lambda_2$ is,

$$\begin{aligned} \frac{1}{N^8} & \left\{ 28 [(N-1)^6 + N-1] [(N-1)^2 + N-1] + (N-1) \left(12 [(N-1)^4 \right. \right. \\ & \quad \left. \left. + (N-1)^2 + N-2] [(N-1)^2 + N-1] + 16 [-2 (N-1)^3 \right. \right. \\ & \quad \left. \left. + N-2] [-2 (N-1) + N-2] \right) \right\} \\ & = \frac{4 N (N-1) (7 N^4 - 39 N^3 + 90 N^2 - 99 N + 45)}{N^7}. \end{aligned}$$

Of $\lambda_5 \lambda_3$

$$\begin{aligned} \frac{1}{N^8} & \left\{ 56 [(N-1)^5 - N+1] [(N-1)^3 - (N-1)] + (N-1) \left(8 [-(N-1)^4 \right. \right. \\ & \quad \left. \left. + (N-1) - (N-2)] [(N-1)^3 - (N-1)] + 48 [(N-1)^3 - (N-1)^2 \right. \right. \\ & \quad \left. \left. - (N-2)] [-(N-1)^2 + (N-1) - (N-2)] \right) \right\} \\ & = \frac{48 N (N-1) (N-2)^2 (N^2 - 3 N + 3)}{N^7}. \end{aligned}$$

And in the same manner the coefficient of λ^2_4 is found to be,

$$\frac{2 N (N-1) (17 N^4 - 111 N^3 + 309 N^2 - 405 N + 207)}{N^7}$$

The coefficient of $\lambda_4 \lambda^2_2$ is,

$$\frac{12 N^2 (N-1) (17 N^3 - 71 N^2 + 117 N - 69)}{N^7}$$

Of $\lambda^2_3 \lambda_2$:

$$\frac{72 N^2 (N-1) (N-2)^2 (3 N - 5)}{N^7}$$

And finally of λ^4_2

$$\frac{24 N^3 (N-1) (4 N^2 - 9 N + 6)}{N^7}$$

Collecting, we have,

$$\begin{aligned} S_2(v_4) = \frac{1}{N^7} & \left[(N-1)^2 (N^2 - 3 N + 3)^2 \lambda_8 + 4 (N-1)^2 (7 N^4 - 39 N^3 \right. \\ & \quad \left. + 90 N^2 - 99 N + 45) \lambda_6 \lambda_2 + 48 N (N-1) (N-2)^2 (N^2 - 3 N + 3) \lambda_5 \lambda_3 \right. \\ & \quad \left. + 2 N (N-1) (17 N^4 - 111 N^3 + 309 N^2 - 405 N + 207) \lambda^2_4 \right. \\ & \quad \left. + 12 N^2 (N-1) (17 N^3 - 71 N^2 + 117 N - 69) \lambda_4 \lambda^2_2 + 72 N^2 (N-1) \right. \\ & \quad \left. (N-2)^2 (3 N - 5) \lambda^2_3 \lambda_2 + 24 N^3 (N-1) (4 N^2 - 9 N + 6) \lambda^4_2 \right]. \end{aligned}$$

There are two main steps in the calculation, the picking out of the coefficients of certain power products of t 's in symbolic expansions of the type $[(\sum \lambda_i t_i)^{(a)}]^r [(\sum \lambda_i t_i)^{(b)}]^s \dots$, and the substitution and reduction of terms of the type $\lambda_{xyz} \dots$ in terms of polynomials in N . The first step can be mechanically carried out by the use of the D -operator of combinatory analysis ⁽¹⁾. An example of its use in this connection follows. It is required to find the coefficient of $t_1^6 t_2^3$ in

$$378 (\sum \lambda_i t_i)^{(6)} [(\sum \lambda_i t_i)^{(2)}]^2.$$

The operators, D_5 , D_2 , and $\bar{D}_2$, are applied in succession on (63), thus,

$$\begin{aligned} D_5 D_2^2 (63) &= D_2^2 [(5) (13) + (41) (22) + (32) (31) + (23) (4)] \\ &= 2 (5) (2) (11) + 2 (41) (2)^2 + (41) (11)^2 + 2 (32) (2) (11) + (32) (2)^2. \end{aligned}$$

In Macmahon's notation this would be written,

$$D_5 D_2^2 (\overline{X_6} \overline{X_3}) = D_2^2 [x_5 \overline{X_1} \overline{X_3} + x_4 x_1 \overline{X_2} \overline{X_2} + x_3 x_2 \overline{X_3} \overline{X_1} + x_2 x_3 \overline{X_4}], \text{ etc.}$$

In the parenthesis notation I use instead, I always keep the partitions of 5, 2, and 2 separated; if I did not but ran them together as MACMAHON does (which is quite satisfactory for his purpose) there would often be ambiguity in the separation which would have to be effected later.

The result indicates the manner in which the total coefficient of t_1^6, t_2^3 , is made up by products of terms from each of the three factors. The first says that there are two made up of a fifth power from the first, a square from one of the last two, and a second degree cross-product from the last two. The terms are actually $t_1^5 \cdot t_2^2 \cdot t_1 t_2$ and $t_1^5 \cdot t_1 t_2 \cdot t_2^2$. This takes no account of the numerical factors (products of binomial coefficients) which would appear in the actual expansion. Supplying them; they are 2 in the first term since $t_1 t_2$ would occur multiplied by 2, 5 in the second since $t_4 t_2$ would enter with a coefficient 5, 20 in the third, and so on. Then writing in the corresponding λ 's, the sought coefficient is,

$$378 (4 \lambda_{50} \lambda_{20} \lambda_{11} + 10 \lambda_{41} \lambda_{20}^2 + 20 \lambda_{41} \lambda_{11}^2 + 40 \lambda_{32} \lambda_{20} \lambda_{11} + 10 \lambda_{32} \lambda_{20}^2)$$

The total of the numerical coefficient inside the parenthesis is 84, which is $\binom{9}{3}$ as it should be, which serves as a check on the calcula-

⁽¹⁾ See MACMAHON, P. A., *An introduction to Combinatory Analysis*; Cambridge University Press, 1920; p. 24 ff.

tion. Obviously the total number of ways of forming terms in $t_1^6 t_2^3$ in

$$(\sum \lambda_i t_i)^{(5)} [(\lambda_i t_i)^{(2)}]^2$$

is the same as if the expression were,

$$(\sum \lambda_i t_i)^{(9)}.$$

But the expression

$$378 (\sum \lambda_i t_i)^{(5)} [(\sum \lambda_i t_i)^{(2)}]^2$$

is but one of the terms in the value of,

$$(\sum \nu_i t_i)^{(9)}$$

in terms of semi-invariants. If I am finding the value of $\nu_{630 \dots 0}$, I recall that this is accomplished by equating coefficients of $t_1^6 t_2^3$ on both sides of the equation which gives

$$(\sum \nu_i t_i)^{(9)}$$

in terms of semi-invariants. It will then be necessary to divide both sides of the equation obtained by thus equating coefficients by the coefficient of $t_1^6 t_2^3$ on the left-hand side of the equation which is, of course, $\binom{9}{3}$. Performing this division the term in the value of $\nu_{630 \dots 0}$ contributed by,

$$378 (\sum \lambda_i t_i)^{(5)} [(\sum \lambda_i t_i)^{(2)}]^2$$

is

$$18 \lambda_{50} \lambda_{20} \lambda_{11} + 45 \lambda_{41} \lambda_{20}^2 + 90 \lambda_{41} \lambda_{11}^2 + 180 \lambda_{32} \lambda_{20} \lambda_{11} + 45 \lambda_{32} \lambda_{20}^2.$$

It is to be noticed that the sum of the numerical coefficients is, as it should be, 378.

CHAPTER IV

The Semi-invariants of the Correlation Function of Two Moments about the Mean

The method which has been developed for finding the semi-invariants of the frequency function for a single moment about the mean is readily extended to finding the semi-invariants of the frequency or correlation function for two moments about the mean. Just as before I find a general formula which gives these semi-invariants in terms of the semi-invariants of $F(\delta_1, \delta_2, \dots, \delta_N)$; once

this is accomplished the calculation of these semi-invariants is carried out in exactly the same way as in the case of a single moment about the mean.

Consider the m th and n th moments about the mean, — the defining equation for the semi-invariants, S_{kl} (v_m , v_n) of the correlation function for them is,

$$e^{(S_{10}s + S_{01}t) + \frac{1}{2!}(S_{10}s + S_{01}t)^{(2)} + \frac{1}{3!}(S_{10}s + S_{01}t)^{(3)} + \dots} \quad (29)$$

$$= \int_{-\infty}^{\infty} d\delta_1 \int_{-\infty}^{\infty} d\delta_2 \dots \int_{-\infty}^{\infty} d\delta_N F(\delta_1, \delta_2, \dots, \delta_N) e^{\frac{(\sum_{i=1}^N \delta_i^m s + \sum_{i=1}^N \delta_i^n t)}{N}}$$

And analogous to the L 's in the case of one moment, now I have L 's and M 's defined by,

$$e^{(\sum_i L_i s_i + \sum_j M_j t_j) + \frac{1}{2!}(\sum_i L_i s_i + \sum_j M_j t_j)^{(2)} + \frac{1}{3!}(\sum_i L_i s_i + \sum_j M_j t_j)^{(3)} + \dots}$$

$$= \int d\delta_1 \dots \int d\delta_N F(\delta_1, \dots, \delta_N) e^{(\sum_i \delta_i^m s_i + \sum_j \delta_j^n t_j)} \quad (30)$$

If in (29) and (30) $n = 0$ the process of finding the L 's and then the S 's is exactly that considered in the last chapter. Thus the S_{i0} 's and the S_{0j} 's, the marginal semi-invariants, have already been obtained and the object now is to learn how to compute the S_{ij} 's, the cross semi-invariants.

First, in analogy with the relation between the S 's and L 's for a single moment, the corresponding relation here is obtained from,

$$S_{h-r, r} s^{h-r} t^r = (\sum_i L_i s_i)^{(h-r)} (\sum_j M_j t_j)^{(r)} / N^h$$

which gives

$$S_{h-r, r} = \frac{1}{N^h} \sum \frac{(h-r)!}{k_1! k_2! \dots k_p!} L_{k_1 k_2 \dots k_p} \sum \frac{r!}{l_1! l_2! \dots l_q!} M_{l_1 l_2 \dots l_q} \quad (31)$$

the summation being performed for all partitions, $k_1, k_2, \dots, k_p$, and $l_1, l_2, \dots, l_q$, of $h-r$ and r respectively.

For the first four orders of cross semi-invariants the full expressions are given (taking into account the fact that L 's and M 's of the same type, respectively, are equal).

$$S_{11} = \frac{1}{N^2} \sum L_{10 \dots 0} \sum M_{01 \dots 0} = \frac{1}{N} L_{10 \dots 0} M_{10 \dots 0} + \frac{(N-1)}{N} L_{10 \dots 0} M_{010 \dots 0} \quad (32)$$

$$\begin{aligned}
S_{21} &= \frac{1}{N^3} (\Sigma L_{20} \dots 0 + 2 \Sigma L_{110} \dots 0) (\Sigma M_{10} \dots 0) \\
&= \frac{1}{N^2} L_{20} M_{10} + \frac{N-1}{N^2} L_{20} M_{01} + \frac{2(N-1)}{N^2} L_{110} M_{10} \\
&\quad + \frac{(N-1)(N-2)}{N^2} L_{110} M_{001} \\
S_{12} &= \frac{1}{N^2} L_{10} M_{20} + \frac{N-1}{N^2} L_{01} M_{20} + \frac{2(N-1)}{N^2} L_{10} M_{110} \\
&\quad + \frac{(N-1)(N-2)}{N^2} L_{001} M_{110} \\
S_{31} &= \frac{1}{N^4} (\Sigma L_{30} + 3 \Sigma L_{210} + 6 \Sigma L_{1110}) (\Sigma M_{10}) \\
&= \frac{1}{N^3} L_{30} M_{10} + \frac{N-1}{N^3} L_{30} M_{01} + \frac{3(N-1)}{N^3} L_{21} M_{10} + \frac{3(N-1)}{N^3} L_{21} M_{01} \\
&\quad + \frac{3(N-1)(N-2)}{N^3} L_{21} M_{001} + \frac{3(N-1)(N-2)}{N^3} L_{1110} M_{10} \\
&\quad + \frac{(N-1)(N-2)(N-3)}{N^3} L_{1110} M_{0001} \quad (32) \\
S_{22} &= \frac{1}{N^4} (\Sigma L_{20} + 2 \Sigma L_{110}) (\Sigma M_{20} + 2 \Sigma M_{110}) \\
&= \frac{1}{N^3} L_{20} M_{20} + \frac{N-1}{N^3} L_{20} M_{02} + \frac{2(N-1)}{N^3} L_{20} M_{110} \\
&\quad + \frac{(N-1)(N-2)}{N^3} L_{20} M_{0110} + \frac{2(N-1)}{N^3} L_{110} M_{20} \\
&\quad + \frac{(N-1)(N-2)}{N^3} L_{0110} M_{20} + \frac{2(N-1)}{N^3} L_{110} M_{110} \\
&\quad + \frac{4(N-1)(N-2)}{N^3} L_{110} M_{011} + \frac{(N-1)(N-2)(N-3)}{N^3} L_{110} M_{00110} . \\
S_{13} &= \frac{1}{N^3} L_{10} M_{30} + \frac{N-1}{N^3} L_{10} M_{03} + \frac{3(N-1)}{N^3} L_{10} M_{210} + \frac{3(N-1)}{N^3} L_{01} M_{210} \\
&\quad + \frac{3(N-1)(N-2)}{N^3} L_{10} M_{1110} + \frac{3(N-1)(N-2)}{N^3} L_{001} M_{210} \\
&\quad + \frac{(N-1)(N-2)(N-3)}{N^3} L_{00010} M_{1110} .
\end{aligned}$$

The next point to be considered is the development of terms of the form $L_{k_1 k_2 \dots k_p}$, $M_{l_1 l_2 \dots l_q}$ into terms of the moments of the correlation function $F(\delta_1, \dots, \delta_N)$. It is obvious from the defining

equation (30) that the general equation (11) applies here if in (11) λ_n be replaced by,

$$(\sum_i L_i s_i + \sum_j M_j t_j)^{(n)}$$

and v_r by

$$(\sum_i v_{mi} s_i + \sum_j v_{nj} t_j)^{(r)}$$

The expansion of the right-hand member of (30) is indicated by

$$1 + (\sum_i v_{mi} s_i + \sum_j v_{nj} t_j) + \frac{1}{2!} (\sum_i v_{mi} s_i + \sum_j v_{nj} t_j)^{(2)} + \frac{1}{3!} (\sum_i v_{mi} s_i + \sum_j v_{nj} t_j)^{(3)} + \dots$$

in which, in the expansion of the various terms, a certain addition of subscripts must take place in the cross-products. This is perhaps best seen from an example. For instance, the expansion of

$$(\sum_i v_{mi} s_i + \sum_j v_{nj} t_j)^{(2)}$$

is

$$\begin{aligned} \sum v_{2m} s_i^2 + \sum v_{m, m} s_i s_j + \sum v_{m+n} s_i t_i + \sum v_m v_n s_i t_j \\ + \sum v_{2n} t_i^2 + \sum v_{n, n} t_i t_j. \end{aligned} \quad (33)$$

As a more extended example of the kind of calculation that is required here consider,

$$\begin{aligned} (\sum_i L_i s_i + \sum_j M_j t_j)^{(3)} = (\sum_i v_{mi} s_i + \sum_j v_{nj} t_j)^{(3)} - 3 (\sum_i v_{mi} s_i + \sum_j v_{nj} t_j)^{(2)} (\sum_i v_{mi} s_i + \sum_j v_{nj} t_j) \\ + 2 (\sum_i v_{mi} s_i + \sum_j v_{nj} t_j)^3. \end{aligned} \quad (34)$$

which comes from (30) and which is analogous to (24). Suppose it is required to find $S_{21} (v_m, v_n)$. First from (32),

$$\begin{aligned} S_{21} = \frac{1}{N^2} L_{20} M_{10} + \frac{N-1}{N^2} L_{20} M_{01} + \frac{2(N-1)}{N^2} L_{110} M_{10} + \\ + \frac{(N-1)(N-2)}{N^2} L_{110} M_{001}. \end{aligned}$$

Then from (34) is obtained (analogous to the second equation of (24a)),

$$\begin{aligned} (\sum L_i s_i)^{(2)} (\sum M_j t_j) = \left\{ (\sum v_{mi} s_i)^{(2)} (\sum v_{nj} t_j) \right\} - (\sum v_{ni} s_i)^{(2)} (\sum v_{nj} t_j) \\ - 2 \left\{ (\sum v_{mi} s_i) (\sum v_{nj} t_j) \right\} (\sum v_{mi} s_i) + 2 (\sum v_{mi} s_i)^2 (\sum v_{nj} t_j) \end{aligned} \quad (35)$$

in which the braces indicate that the terms in each pair are from the same symbolic expansion term of (34) and that hence the addition of subscripts mentioned above must be performed therein. The ex-

pression for S_{21} obtained from (32) shows the types of terms to be found from (35), viz., $L_{20} M_{10}$, $L_{20} M_{01}$, $L_{110} M_{10}$, and $L_{110} M_{001}$. To get these simply requires the coefficients of $s_1^2 t_1$, $s_1^2 t_2$, $s_1 s_2 t_1$, and $s_1 s_2 t_3$ in the right-hand member of (35) respectively. Writing these out, they are

$$\begin{aligned} L_{20} M_{10} &= v_{2m+n,0} - v_{2m,0} v_{n,0} - 2 v_{m+n,0} v_{m,0} + 2 v_{m,0}^2 v_{n,0} \\ L_{20} M_{01} &= v_{2m,n,0} - v_{2m,0} v_{n,0} - 2 v_{m,n,0} v_{m,0} + 2 v_{m,0}^2 v_{n,0} \\ L_{110} M_{10} &= v_{m+n,m,0} - v_{m,m,0} v_{n,0} - v_{m+n,0} v_{m,0} - v_{m,n,0} v_{m,0} + 2 v_{m,0}^2 v_{n,0} \\ L_{110} M_{001} &= v_{m,m,n,0} - v_{m,m,0} v_{0,0,n,0} - 2 v_{m,n,0} v_{m,0} + 2 v_{m,0}^2 v_{n,0} \end{aligned} \quad (36)$$

(In the third term of the last equation two terms of the same type are combined). Substituting these values in S_{21} and collecting,

$$\begin{aligned} S_{21} = \frac{1}{N^2} & \left[v_{2m+n,0} + (N-1) v_{2m,n,0} + 2(N-1) v_{m+n,n,0} \right. \\ & + (N-1)(N-2) v_{m,m,n,0} - N v_{2m,0} v_{n,0} - N(N-1) v_{m,m,0} v_{n,0} \\ & \left. - 2N v_{m+n,0} v_{m,0} - 2N(N-1) v_{m,n,0} v_{m,0} + 2N^2 v_{m,0}^2 v_{n,0} \right]. \end{aligned} \quad (37)$$

As further illustrations the calculation of two of the simplest cases, $S_{11}(v_2, v_3)$ and $S_{11}(v_2, v_4)$ are shown in some detail.

First, in general, from (32),

$$S_{11} = \frac{1}{N} L_{10} M_{10} + \frac{N-1}{N} L_{10} M_{01}.$$

for any two moments, v_m and v_n . Then from the relation connecting semi-invariants and moments in general,

$$(\sum L_i s_i + \sum M_j t_j)^{(2)} = (\sum v_{2i} s_i + \sum v_{3j} t_j)^{(2)} - (\sum v_{2i} s_i + \sum v_{3j} t_j)^2.$$

This gives

$$(\sum L_i s_i)(\sum M_j t_j) = \left\{ (\sum v_{2i} s_i)(\sum v_{3j} t_j) \right\} - (\sum v_{2i} s_i)(\sum v_{3j} t_j)$$

from which

$$\begin{aligned} L_{10} M_{10} &= v_{50} \dots 0 - v_{20} \dots 0 v_{30} \dots 0 \\ L_{10} M_{01} &= v_{320} \dots 0 - v_{20} \dots 0 v_{30} \dots 0 \end{aligned}$$

Substituting these in (38),

$$\begin{aligned} S_{11}(v_2, v_3) &= \frac{1}{N} \left[v_{50} - v_{20} v_{30} + (N-1)(v_{32} - v_{20} v_{30}) \right] \\ &= \frac{1}{N} \left[v_{50} + (N-1)v_{32} - N v_{20} v_{30} \right]. \end{aligned} \quad (39)$$

From

$$(\sum v_i t_i)^{(5)} = (\sum \lambda_i t_i)^{(5)} + 10 (\sum \lambda_i t_i)^{(3)} (\sum \lambda_i t_i)^{(2)},$$

are obtained,

$$\begin{aligned} v_{50} &= \lambda_{50} + 10 \lambda_{20} \lambda_{30} \\ v_{32} &= v_{32} + \lambda_{30} \lambda_{20} + 6 \lambda_{21} \lambda_{11} + 3 \lambda_{21} \lambda_{20}, \end{aligned}$$

and setting these in (39),

$$\begin{aligned} S_{11}(v_2, v_3) &= \frac{1}{N} \left[\lambda_{50} + 10 \lambda_{30} \lambda_{20} + (N-1)(\lambda_{32} + \lambda_{30} \lambda_{20} + 6 \lambda_{21} \lambda_{11} \right. \\ &\quad \left. + 3 \lambda_{21} \lambda_{20} - N \lambda_{30} \lambda_{20}) \right] \\ &= \frac{1}{N} \left[\lambda_{50} + (N-1) \lambda_{32} + 9 \lambda_{30} \lambda_{20} + (N-1)(6 \lambda_{21} \lambda_{11} + 2 \lambda_{21} \lambda_{20}) \right] \end{aligned}$$

And on replacing the semi-invariants of $F(\delta_1, \dots, \delta_N)$ by their values in terms of the semi-invariants of $f(x)$, finally,

$$\begin{aligned} S_{11}(v_2, v_3) &= \frac{1}{N^6} \left[N^2 (N-1)^2 (N-2) \lambda_5 + 6 N^3 (N-1) (N-2) \lambda_3 \lambda_2 \right] \\ &= \frac{1}{N^4} \left[(N-1)^2 (N-2) \lambda_5 + 6 N (N-1) (N-2) \lambda_3 \lambda_2 \right] \quad (40) \end{aligned}$$

For $S_{11}(v_2, v_4)$, analogous to (39),

$$S_{11}(v_2, v_4) = \frac{1}{N} \left[v_{60} + (N-1) v_{42} - N v_{40} v_{20} \right] \quad (41)$$

From

$$\begin{aligned} (\Sigma v_i t_i)^{(6)} &= (\Sigma \lambda_i t_i)^{(6)} + 15 (\Sigma \lambda_i t_i)^{(4)} (\Sigma \lambda_i t_i)^{(2)} + \\ &\quad + 10 [(\Sigma \lambda_i t_i)^{(3)}]^2 + 15 [(\Sigma \lambda_i t_i)^{(2)}]^3. \end{aligned}$$

are obtained

$$\begin{aligned} v_{60} &= \lambda_{60} + 15 \lambda_{40} \lambda_{20} + 10 \lambda_{30}^2 + 15 \lambda_{20}^3 \\ v_{42} &= \lambda_{42} + (\lambda_{40} \lambda_{20} + 8 \lambda_{31} \lambda_{11} + 6 \lambda_{22} \lambda_{20}) + (4 \lambda_{30} \lambda_{21} + 6 \lambda_{21}^2) + (3 \lambda_{20}^3 + 12 \lambda_{20} \lambda_{11}^2) \end{aligned}$$

Also

$$v_{40} v_{20} = \lambda_{40} \lambda_{20} + 3 \lambda_{20}^3.$$

Substituting these in (41), collecting and arranging terms,

$$\begin{aligned} S_{11}(v_2, v_4) &= \frac{1}{N} \left[\lambda_{60} + (N-1) \lambda_{42} + 14 \lambda_{40} \lambda_{20} + (N-1)(8 \lambda_{31} \lambda_{11} + 6 \lambda_{22} \lambda_{20}) \right. \\ &\quad \left. + 10 \lambda_{30}^2 + (N-1)(4 \lambda_{30} \lambda_{21} + 6 \lambda_{21}^2) + 12 \lambda_{20}^3 + 12 (N-1) \lambda_{20} \lambda_{11}^2 \right] \end{aligned}$$

Again introducing the semi-invariants of $f(x)$, reducing and cancelling an N^2 from numerator and denominator, finally

$$\begin{aligned} S_{11}(v_2, v_4) &= \frac{1}{N^5} \left[(N-1)^2 (N^2 - 3N + 3) \lambda_6 + 2N(N-1)(7N^2 - 18N + 15) \lambda_4 \lambda_2 \right. \\ &\quad \left. + 6N(N-1)(N-2)^2 \lambda_3^2 + 12N^2(N-1)^2 \lambda_3^2 \right] \quad (22) \end{aligned}$$

It may be observed by way of application that if in these results, $S_{11}(\nu_2, \nu_3)$ were divided by $\sqrt{S_2(\nu_2)} \cdot \sqrt{S_2(\nu_3)}$ and

$$S_{11}(\nu_2, \nu_4) \text{ by } \sqrt{S_2(\nu_2)} \cdot \sqrt{S_2(\nu_4)}$$

the results would be the ordinary Pearsonian coefficients of correlation between ν_2 and ν_3 , and ν_2 and ν_4 , respectively. If, in particular, $f(x)$ be assumed normal, in which case $\lambda_3 = \lambda_4 = \lambda_5 = \dots = 0$, then

$$r_{\nu_2, \nu_3} = 0$$

and

$$r_{\nu_2, \nu_4} = \frac{N - 1}{2\sqrt{4N^2 - 9N + 6}}.$$

It is natural to look for a general law for writing out S_{pq} 's in terms of the moments of $F(\delta_1, \dots, \delta_N)$, analogous to the one given for writing out S_p 's given by (25). This latter should, of course, be a special case of the law for S_{pq} 's, found by setting $q = 0$. Its form is easily obtained by analogy. In addition to (37) I also worked out in detail $S_{31}(\nu_m, \nu_n)$. This further confirmed that the law in question is

$$S_{kl}(\nu_m, \nu_n) = \frac{1}{N^{k+l}} \times \quad (43)$$

$$\sum \frac{(-1)^{(r+s+t+\dots)-1} [(r+s+t+\dots)-1]! k! l! \nu_{a_1 n, a_2 n, \dots, \nu_{a_1 m, a_2 m, \dots} \{^r \nu_{b_1 n, b_2 n, \dots, \nu_{\beta_1 n, \beta_2 n, \dots} \{^s \dots}}{(a_1! a_2! \dots a_1! a_2! \dots)^r (b_1! b_2! \dots \beta_1! \beta_2! \dots)^s \dots r! s!}$$

in which

$$\begin{aligned} r(a_1 + a_2 + \dots) + s(b_1 + b_2 + \dots) + \dots &= k \\ r(\alpha_1 + \alpha_2 + \dots) + s(\beta_1 + \beta_2 + \dots) + \dots &= l \\ a_1 \geq a_2 \geq \dots & \\ b_1 \geq b_2 \geq \dots & \\ \dots & \\ \alpha_1 \geq \alpha_2 \geq \dots & \\ \beta_1 \geq \beta_2 \geq \dots & \\ \dots & \\ (a_1 + a_2 + \dots) > (b_1 + b_2 + \dots) > \dots & \\ (\alpha_1 + \alpha_2 + \dots) > (\beta_1 + \beta_2 + \dots) > \dots & \end{aligned}$$

This law may readily be derived directly. Going back to (29), it is seen that,

$$\begin{aligned} (S_{10}s + S_{01}t)^{(k+l)} &= \\ \frac{1}{N^{k+l}} \sum \frac{(-1)^{(r+s+t+\dots)-1} [(r+s+t+\dots)-1]! (k+l)!}{(a!)^r (b!)^s \dots r! s! \dots} \times \\ &\times [(\sum \nu_{mi} s_i + \sum \nu_{nj} t_j)^{(a)}]^r [(\sum \nu_{mi} s_i + \sum \nu_{nj} t_j)^{(b)}]^s \dots \end{aligned}$$

in which,

$$ar + bs + \dots = k + l,$$

etc., (see under (10)).

To find S_{kl} , observe that on the left-hand side the coefficient of $s^k t^l$ is $\frac{(k+l)!}{k! l!}$ and that this is equal to on the right-hand side the sum of the coefficients of all such terms as

$$s_1^{a_1 r} s_2^{a_2 r} \dots s_1^{b_1 s} s_2^{b_2 s} \dots t_1^{a_1 r} t_2^{a_2 r} \dots t_1^{b_1 s} t_2^{b_2 s} \dots$$

in which

$$r(a_1 + a_2 + \dots) + s(b_1 + b_2 + \dots) + \dots = k$$

$$r(\alpha_1 + \alpha_2 + \dots) + s(\beta_1 + \beta_2 + \dots) + \dots = l.$$

This sum divided by $\frac{(k+l)!}{k! l!}$ gives then, the value for S_{kl} and this is precisely the right-hand side of (43).

As a conclusion to this section I have here collected the general expressions for $S_{kl}(\nu_m, \nu_n)$ up to the fourth order, some of which are already derived above directly.

$$S_{11}(\nu_m, \nu_n) = \frac{1}{N} \left[\nu_{m+n,0} + (N-1) \nu_{m,n,0} - N \nu_{m,0} \nu_{n,0} \right]$$

$$\begin{aligned} S_{21}(\nu_m, \nu_n) = \frac{1}{N^2} & \left[\nu_{2m+n,0} + (N-1) \nu_{2m,n,0} + 2(N-1) \nu_{m+n,m,0} \right. \\ & + (N-1)(N-2) \nu_{m,m,n,0} - N \nu_{2m,0} \nu_{n,0} - N(N-1) \nu_{m,m,0} \nu_{n,0} \\ & \left. - 2N \nu_{m+n,0} \nu_{m,0} - 2N(N-1) \nu_{m,n,0} \nu_{m,0} + 2N^2 \nu_{m,0}^2 \nu_{n,0} \right]. \end{aligned}$$

$S_{12}(\nu_m, \nu_n)$ is, of course, the above with the m and n interchanged.

$$\begin{aligned} S_{31}(\nu_m, \nu_n) = \frac{1}{N^3} & \left[\nu_{3m+n,0} + (N-1) \nu_{3m,n,0} + 3(N-1) \nu_{2m+n,m,0} \right. \\ & + 3(N-1) \nu_{2m,m+n,0} + 3(N-1)(N-2) \nu_{2m,m,n,0} \\ & + 3(N-1)(N-2) \nu_{m+n,m,m,0} + (N-1)(N-2)(N-3) \nu_{m,m,m,n,0} \\ & - N \nu_{3m,0} \nu_{n,0} - 3N(N-1) \nu_{2m,m,0} \nu_{n,0} \\ & - N(N-1)(N-2) \nu_{m,m,m,0} \nu_{n,0} - 3N \nu_{2m+n,0} \nu_{m,0} \\ & - 3N(N-1) \nu_{2m,n,0} \nu_{m,0} - 6N(N-1) \nu_{m+n,m,0} \nu_{m,0} \\ & - 3N(N-1)(N-2) \nu_{m,m,n,0} \nu_{m,0} - 3N \nu_{2m,0} \nu_{m+n,0} \\ & - 3N(N-1) \nu_{2m,0} \nu_{m,n,0} - 3N(N-1) \nu_{m,m,0} \nu_{m+n,0} \\ & - 3N(N-1)^2 \nu_{m,m,0} \nu_{m,n,0} + 6N^2 \nu_{2m,0} \nu_{m,0} \nu_{n,0} + 6N^2 \nu_{m+n,0} \nu_{m,0}^2 \\ & + 6N^2(N-1) \nu_{m,m,0} \nu_{m,0} \nu_{n,0} + 6N^2(N-1) \nu_{m,n,0} \nu_{m,0}^2 \\ & \left. - 6N^3 \nu_{m,0}^3 \nu_{n,0} \right] \end{aligned} \quad (45)$$

$$\begin{aligned}
S_{22}(\nu_m, \nu_n) = \frac{1}{N^3} \bigg[& \nu_{2m+2n,0} + 2(N-1)\nu_{2m+n,n,0} + 2(N-1)\nu_{m+2n,m,0} \\
& + (N-1)\nu_{2m,2n,0} + 2(N-1)\nu_{m+n,m+n,0} + (N-1)(N-2)\nu_{2m,n,n,0} \\
& + (N-1)(N-2)\nu_{2n,m,m,0} + 4(N-1)(N-2)\nu_{m+n,m,n,0} \\
& + (N-1)(N-2)(N-3)\nu_{m,m,n,n,0} - 2N\nu_{2m+n,0}\nu_{n,0} \\
& - 2N\nu_{2n+m,0}\nu_{m,0} - 2N(N-1)\nu_{2m,n,0}\nu_{n,0} - 2N(N-1)\nu_{2n,m,0}\nu_{m,0} \\
& - 4N(N-1)\nu_{m+n,m,0}\nu_{n,0} - 4N(N-1)\nu_{m+n,n,0}\nu_{m,0} \quad (45) \\
& - 2N(N-1)(N-2)\nu_{m,m,n,0}\nu_{n,0} - 2N(N-1)(N-2)\nu_{n,n,m,0}\nu_{m,0} \\
& - N\nu_{2m,0}\nu_{2n,0} - N(N-1)\nu_{2m,0}\nu_{n,n,0} - N(N-1)\nu_{2n,0}\nu_{m,m,0} \\
& - 2N\nu_{m+n,0}^2 - 4N(N-1)\nu_{m+n,0}\nu_{m,n,0} - 2N(N-1)^2\nu_{m,n,0}^2 \\
& - N(N-1)^2\nu_{m,m,0}\nu_{n,n,0} + 2N^2\nu_{2m,0}\nu_{n,0}^2 + 2N^2\nu_{2n,0}\nu_{m,0}^2 \\
& + 8N^2\nu_{m+n,0}\nu_{m,0}\nu_{n,0} + 8N^2(N-1)\nu_{m,n,0}\nu_{m,0}\nu_{n,0} \\
& + 2N^2(N-1)\nu_{m,0}^2\nu_{n,n,0} + 2N^2(N-1)\nu_{n,0}^2\nu_{m,m,0} - 6N^3\nu_{m,0}^2\nu_{n,0}^2 \bigg]
\end{aligned}$$

$S_{13}(\nu_m, \nu_n)$ is obtained from $S_{31}(\nu_m, \nu_n)$.

The form of the result (44) suggests the generalization which could be effected for finding the semi-invariants of the correlation function for any number of moments about the mean. But the applications to be made in the present paper only require the results for two moments and nothing would be gained by writing out the even more complicated expressions analogous to those of (45).

CHAPTER V

The Correlation Function for Powers of Moments

It is of some interest and it is also a further illustration of the method of semi-invariants to show how use can be made of the results of the two preceding chapters in finding the semi-invariants of powers of moments about the mean, that is, of the frequency functions of such moments.

A. Semi-invariants of Powers of Moments about the Mean.

After the semi-invariants of the frequency function of ν_n have been found, the semi-invariants of ν_n^p are found in exactly the same way that the semi-invariants of the frequency function for x^p are found from the semi-invariants of $f(x)$, the frequency function for

x . It is to be recalled that the semi-invariants of $f(x)$ are defined by,

$$e^{\lambda_1 t + \frac{1}{2!} \lambda_2 t^2 + \frac{1}{3!} \lambda_3 t^3 + \dots} = \int_{-\infty}^{\infty} dx f(x) e^{xt} \quad (46)$$

Then the semi-invariants, L_i , for the distribution function for x^p are defined by,

$$e^{L_1 t + \frac{1}{2!} L_2 t^2 + \frac{1}{3!} L_3 t^3 + \dots} = \int_{-\infty}^{\infty} dx f(x) e^{x^p t} \quad (47)$$

On expansion of the right-hand member it appears from comparison with (11) that L_i can be written out in terms of the moments of $f(x)$ by means of the relation,

$$L_n = \Sigma \frac{(-1)^{(r+s+t+\dots)-1} [(r+s+t+\dots)-1]! n! v_{pa}^r v_{pb}^s v_{pc}^t \dots}{(a!)^r (b!)^s (c!)^t \dots r! s! t! \dots} \quad (48)$$

with the same restrictions on $a, b, c, \dots$, and $r, s, t, \dots$, as in (10). But since by (10) the moments of $f(x)$ can be immediately written out in terms of the semi-invariants of $f(x)$, the L 's can readily be calculated in terms of λ 's.

As an example, let. $p = 2$ and the origin be at the mean; to find L_1, L_2, L_3 , and L_4 . By (48),

$$\begin{aligned} L_1 &= v_2 \\ L_2 &= v_4 - v_2^2 \\ L_3 &= v_6 - 3 v_4 v_2 + 2 v_2^3 \\ L_4 &= v_8 - 4 v_6 v_2 - 3 v_4^2 + 12 v_4 v_2^2 - 6 v_2^4. \end{aligned}$$

And from (10),

$$\begin{aligned} v_2 &= \lambda_2 \\ v_3 &= \lambda_3 \\ v_4 &= \lambda_4 + 3 \lambda_2^2 \\ v_6 &= \lambda_6 + 15 \lambda_4 \lambda_2 + 10 \lambda_3^2 + 15 \lambda_2^3 \\ v_8 &= \lambda_8 + 28 \lambda_6 \lambda_2 + 56 \lambda_5 \lambda_3 + 35 \lambda_4^2 + 210 \lambda_4 \lambda_2^2 + 280 \lambda_3^2 \lambda_2 + 105 \lambda_2^4. \end{aligned}$$

Substituting these values in the above and collecting,

$$\begin{aligned} L_1 &= \lambda_2 \\ L_2 &= \lambda_4 + \lambda_2^2 \\ L_3 &= \lambda_6 + 12 \lambda_4 \lambda_2 + 10 \lambda_3^2 + 8 \lambda_2^3 \\ L_4 &= \lambda_8 + 24 \lambda_6 \lambda_2 + 56 \lambda_5 \lambda_3 + 32 \lambda_4^2 + 144 \lambda_4 \lambda_2^2 + 240 \lambda_3^2 \lambda_2 + 48 \lambda_2^4. \end{aligned}$$

Incidentally it appears that if $f(x)$ is normal the distribution of x^2 is by no means normal.

B. *Semi-invariants of the Correlation Function for Powers of two Moments about the Mean.*

After the semi-invariants of the correlation function for v_m and v_n have been found, the semi-invariants for the correlation function for v_m^p and v_n^q are found in precisely the same way that the semi-invariants of the correlation function for x^p and y^q are found from the semi-invariants of $\varphi(x, y)$, the correlation function for x and y . And it is easy to see how this latter process is carried out. The semi-invariants of $\varphi(x, y)$ are defined from the equation,

$$e^{(\lambda_{10}s + \lambda_{01}t) + \frac{1}{2!}(\lambda_{10}s + \lambda_{01}t)^{(2)} + \frac{1}{3!}(\lambda_{10}s + \lambda_{01}t)^{(3)} + \dots} = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \varphi(x, y) e^{(xs + yt)} \quad (49)$$

The semi-invariants for the correlation function for x^p and y^q are defined by,

$$e^{(L_{10}s + L_{01}t) + \frac{1}{2!}(L_{10}s + L_{01}t)^{(2)} + \frac{1}{3!}(L_{10}s + L_{01}t)^{(3)} + \dots} = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \varphi(x, y) e^{(x^p s + y^q t)} \quad (50)$$

Again after expansion of the right-hand of 50) and comparison with the general expression for semi-invariants in terms of moments, it is seen that,

$$e^{(L_{10}s + L_{01}t)^{(k+l)}} = \sum \frac{(-1)^{(r+s+t+\dots)-1} [(r+s+t+\dots)-1]! (k+l)!}{(a!)^r (b!)^s (c!)^t \dots r! s! t! \dots} \times \frac{[(v_p s + v_q t)^{(a)}]^r [(v_p s + v_q t)^{(b)}]^s \dots}{(51)}$$

in which

$$ar + bs + \dots = k + l, \quad \text{etc.},$$

from which, just as in (43),

$$L_{kl} = \sum \frac{(-1)^{(r+s+t+\dots)-1} [(r+s+t+\dots)-1]! k! l!}{(a_1! a_2! \dots a_1! a_2! \dots)^r (b_1! b_2! \dots \beta_1! \beta_2! \dots)^s \dots} \times \frac{\{v_{a_1 p, a_2 p, \dots, v_{a_1 q, a_2 q, \dots, v_{b_1 p, b_2 p, \dots, v_{\beta_1 q, \beta_2 q, \dots}\}^r \{v_{b_1 p, b_2 p, \dots, v_{\beta_1 q, \beta_2 q, \dots}\}^s \dots}}{r! s! \dots} \quad (52)$$

with the same restrictions on the a 's, b 's, ..., α 's, β 's, ..., r 's, s 's, t 's, ... as in (43). From (49) and the general relation

for moments in terms of semi-invariants,

$$v_{kl} = \Sigma \frac{k! l! [\lambda_{a_1 a_2} \lambda_{a_1 a_2}]^r [\lambda_{b_1 b_2} \lambda_{b_1 b_2}]^s \dots}{(a_1! a_2! \alpha_1! \alpha_2!)^r (b_1! b_2! \beta_1! \beta_2!)^s \dots r! s! \dots} \quad (53)$$

Then by means of the these two last equations, the L_{kl} 's can be computed in terms of the semi-invariants of $\varphi(x, y)$.

CHAPTER VI.

On Obtaining Results to Given Degrees of Approximation.

The third and fourth chapters of this paper describe straightforward and quite mechanical processes for the computation of any semi-invariant of any moment about the mean, or for finding the cross-semi-invariants for any two moments about the mean. The process is reduced to a routine of algebraic computation that could be taught to any intelligent high school graduate. No knowledge of the properties of a special class of functions, such as KRAMP's faculty functions, is needed to carry the calculation of $S_p(v_m)$ or $S_{pq}(v_m, v_n)$ thru in detail.

But it has to be admitted that the systematic calculation of the first semi-invariants of the first four moments about the mean becomes very long and tedious for $S_3(v_4)$ and $S_4(v_3)$, and that for $S_4(v_4)$ truly an enormous amount of labor is involved in getting its complete expression. As it appears from writing out v_{16} in terms of λ 's by means of the relation connecting moments and semi-invariants, the complete expression for $S_4(v_4)$ contains 55 terms, each one a polynomial in N (of degree eight or higher as is apparent from other considerations) times a product, $\lambda_r \lambda_s \lambda_t \dots$, of semi-invariants of $f(x)$ such that $r + s + t + \dots = 16$. I suppose that everybody who has attacked the problem of describing the variation of moments about the mean due to sampling has done so with the hope of discovering a direct and comparatively simple law for writing out the m th moment or semi-invariant of the n th moment about the mean or semi-invariant. One of the most brilliant of the persons to attack this problem was THIELE himself. He wrote out a few of the semi-invariants of semi-invariants but he gave it up as too difficult long before he got to $S_4(\lambda_4)$ ⁽¹⁾. I have carried his method far enough to see that it

⁽¹⁾ THIELE, T. N., Loc. cit., p. 44 ff.

would hardly be possible to take it any further than he did. Certainly the results he obtained gave him no clue at all to any general form of semi-invariants of semi-invariants.

Consider the coefficient of λ_2^8 in $S_4(v_4)$; it appears to be

$$\frac{6912 (N-1) (272 N^4 - 1467 N^3 + 3363 N^2 - 3717 N + 1638)}{N^8} \quad (54)$$

and it is presumably one of the simplest of the 55 in the full expression. It does not look as if study of it would be likely to help much in the discovery of any general law. In fact, when I consider that any general expression for $S_m(v_n)$ must be capable of expansion into 55 terms for m and n each equal to 4 and into a great many more if m and n be each increased by unity, and that it must give complicated and apparently unpredictable polynomials in N for each one of its terms, these facts seem to me to preclude any possibility of a general expression for $S_m(v_n)$ which is at all simple. A high degree of complexity is inherent in the problem and it may even be, fully conceived, that it is of the same order as that of the general expression for $S_m(v_n)$ I have shown the possibility of writing above. Such a conclusion looks reasonable to me after two years study of the matter. It rather seems that the best hopes of effectively further simplifying the problem of sampling for statistical characteristics lie either in the discovery of a new kind of symmetric function of all the observations which may be used to characterize frequency functions and which will be more amenable than either moments or semi-invariants for use in sampling problems, or in, what may very well prove to be much better and more feasible, the abandonment of the method of characterizing frequency functions by symmetric functions of all the observations altogether,

But, on the other hand, it is certainly a pertinent question to inquire just what use could be made of $S_4(v_4)$ if it were written out in its entirety. The answer is, of course, that the magnitude or order of magnitude of the successive terms could be studied. But the method developed in this paper provides the means of determining these without having to write out the complete expressions or even the separate terms in full. If the complete expression were obtained the next step would be to introduce hypotheses which would permit greatly cutting its unwieldy bulk for practical use. Aside from its mere size a more important difficulty arises from the fact that it

would contain in all but the last few terms semi-invariants of orders higher than are ever computed in practice from sample distributions for reasons that are universally recognized. (All but the last ten terms of $S_4(v_4)$ contain semi-invariants of order higher than four). It has been the general practice to follow the advice of THIELE in regard to the use of the higher semi-invariants, namely, that they should be determined from theoretical considerations rather than from computation alone ⁽¹⁾.

The simplest case of such theoretical considerations is the one in which it is supposed that the parent distribution from which the sample is taken is normal, in which case all the semi-invariants of order higher than the second vanish. Another is to assume that in case λ_3 and λ_4 are small that the parent distribution can be described with the three terms of the type A series alone. THIELE gives a warning, hardly needed, against setting the higher semi-invariants equal to zero when λ_3 and λ_4 are not small ⁽²⁾. Still another is explained by WICKSELL and comes naturally from his genetic theory of frequency functions and from the addition theorem for semi-invariants which establishes the second of the important properties of semi-invariants mentioned in the first chapter ⁽³⁾. It is, in particular, that if in the type A expansion for $f(x)$,

$$f(\xi) = \varphi(\xi) + \beta_3 \frac{d^3 \varphi(\xi)}{d\xi^3} + \beta_4 \frac{d^4 \varphi(\xi)}{d\xi^4} + \beta_5 \frac{d^5 \varphi(\xi)}{d\xi^5} \\ + \left(\beta_6 + \frac{\beta_3^2}{2} \right) \frac{d^6 \varphi(\xi)}{d\xi^6} + \dots$$

where

$$\xi = \frac{x - m_x}{\sigma_x}$$

and

$$\beta_r = \frac{\lambda_r}{r! \lambda_2^{r/2}}, \quad (r \geq 3)$$

that the β 's decrease in the constant ratio of $s^{1/2}$ where s is the number of error sources contributing to the value of x . That is, if $\lambda_2^{1/2}$ be taken of order of unity, β_3 will be of order $s^{-1/2}$, β_4 of order s^{-1} , and so on. If further an assumption be made as to the comparative magni-

⁽¹⁾ THIELE, T. N., Loc. cit., p. 49.

⁽²⁾ THIELE, T. N., Loc. cit., p. 49.

⁽³⁾ WICKSELL, S. D., Loc. cit., (on p. 4) p. 16 ff.

tudes of s and N , the number of repetitions composing a sample, the process of calculation here outlined permits the computation of the $S_m(v_n)$'s to any desired degree of approximation without inclusion of superfluous terms in the process, a circumstance that greatly decreases the amount of algebraic labor involved.

As an example of the usefulness in computation of this last assumption, let it be assumed that s and N are both at least equal to 100, and let it be required to approximate the value of $S_3(v_4)/\lambda_2^6$ to the order $1/100$. (Such an assumption as to the magnitude of s and N is immediately applicable in all cases in which $\beta_4 \leq .01$ and $N \geq 100$; and it is to be observed that even so extreme a value of β_4 as .04 is compensated for by an $N \geq 400$, for saving all terms of order $1/N$ and over will still give an approximation of order $1/100$). In the calculation of $S_3(v_4)$ in full the first step is the substitution in the third formula of (27). This is quickly done and the next step is the conversion of the $v_{rst} \dots$'s into semi-invariants $\lambda_{rst} \dots$ of $F(\delta_1, \dots, \delta_N)$ by the use of the expression for moments in terms of semi-invariants and the D-operator as previously explained. This is carried out in order for each of the terms of the final result, all the parts that contribute to each one being obtained in full before passing on to the next. For instance, one term in $S_3(v_4)$ will be a polynomial in N times $\lambda_4 \lambda_2^4$, and all the terms which contribute to this polynomial are obtained independently of and separately from the others. Fixing the attention on this one term, the polynomial times $\lambda_4 \lambda_2^4$ is obtained on the reduction of:

$$\begin{aligned} \frac{1}{N^2} & \left[51975 \lambda_{40} \lambda_{20}^4 + 3(N-1)(735 \lambda_{40} \lambda_{20}^4 + 5040 \lambda_{40} \lambda_{20}^2 \lambda_{11}^2 + 1680 \lambda_{40} \lambda_{11}^4 \right. \\ & + 13440 \lambda_{31} \lambda_{20}^3 \lambda_{11} + 13440 \lambda_{31} \lambda_{20} \lambda_{11}^3 + 2520 \lambda_{22} \lambda_{20}^4 + 15120 \lambda_{22} \lambda_{20}^2 \lambda_{11}^2 \\ & + (N-1)(N-2)(27 \lambda_{40} \lambda_{20}^4 + 216 \lambda_{40} \lambda_{20}^2 \lambda_{11}^2 + 72 \lambda_{40} \lambda_{11}^4 \\ & + 864 \lambda_{31} \lambda_{20}^3 \lambda_{11} + 3456 \lambda_{31} \lambda_{20}^2 \lambda_{11}^2 + 3456 \lambda_{31} \lambda_{20} \lambda_{11}^3 + 2304 \lambda_{31} \lambda_{11}^4 \\ & + 324 \lambda_{22} \lambda_{20}^4 + 3240 \lambda_{22} \lambda_{20}^2 \lambda_{11}^2 + 5184 \lambda_{22} \lambda_{20} \lambda_{11}^3 + 2592 \lambda_{22} \lambda_{11}^4 \\ & + 2592 \lambda_{211} \lambda_{20}^3 \lambda_{11} + 5184 \lambda_{211} \lambda_{20}^2 \lambda_{11}^2 + 12096 \lambda_{211} \lambda_{20} \lambda_{11}^3 + 10368 \lambda_{211} \lambda_{11}^4) \\ & - 3N(105 \lambda_{40} \lambda_{20}^4 + 630 \lambda_{40} \lambda_{20}^2 \lambda_{11}^2) - 3N(N-1)(9 \lambda_{40} \lambda_{20}^4 + 72 \lambda_{40} \lambda_{20}^2 \lambda_{11}^2 \\ & + 24 \lambda_{40} \lambda_{11}^4 + 18 \lambda_{40} \lambda_{20}^4 + 288 \lambda_{31} \lambda_{20}^3 \lambda_{11} + 108 \lambda_{22} \lambda_{20}^4 + 216 \lambda_{22} \lambda_{20}^2 \lambda_{11}^2) \\ & \left. + 54 N^2 \lambda_{40} \lambda_{20}^4 \right] \end{aligned}$$

Noting that,

$$\frac{\lambda_4 \lambda_2^4}{\lambda_2^6} = \beta_4 \cdot 4!$$

it is seen that in the expression just written out only terms of order zero and over need be saved. But it is extremely useful here to note that, except for numerical coefficients, on collection of coefficients of like terms that no term is of order higher than -2 in N . It is necessary now to recall the general expression for terms of the type $\lambda_{rst} \dots$ in terms of semi-invariants of $f(x)$. The equation (17) which gives this is here repeated,

$$\lambda_{k_1 k_2 k_3 \dots k_N} = \lambda_k \frac{1}{N^k} \left[\sum_{i=1}^N (-1)^{k-k_i} (N-1)^{k_i} \right],$$

in which

$$k_1 + k_2 + k_3 + \dots + k_N = k.$$

Thus is obtained,

$$\begin{aligned} \lambda_{40} &= \frac{\lambda_4}{N^4} \left[(N-1)^4 + (N-1) \right] = \frac{N(N-1)(N^2-3N+3)}{N^4} \lambda_4 \\ \lambda_{31} &= \frac{\lambda_4}{N^4} \left[-(N-1)^3 - (N-1) + (N-2) \right] = -\frac{N(N^2-3N+3)}{N^4} \lambda_4 \\ \lambda_{22} &= \frac{\lambda_4}{N^4} \left[2(N-1)^2 + (N-2) \right] = \frac{N(2N-3)}{N^4} \lambda_4 \\ \lambda_{211} &= \frac{\lambda_4}{N^4} \left[(N-1)^2 - 2(N-1) + N-3 \right] = \frac{N(N-3)}{N^4} \lambda_4 \\ \lambda_{20} &= \frac{\lambda_2}{N^2} \left[(N-1)^2 + N-1 \right] = \frac{N(N-1)}{N^2} \lambda_2 \\ \lambda_{11} &= \frac{\lambda_2}{N^2} \left[-2(N-1) + N-2 \right] = -\frac{N}{N^2} \lambda_2, \end{aligned}$$

and in general the value of $\lambda_{k_1 k_2 \dots k_N}$ given directly on substitution in (17) can be written as fast as one can write. The important fact for the approximation process in hand is that $\lambda_{k 0 \dots 0}$ is of order zero in N , but that $\lambda_{k_1 k_2 \dots k_N}$ is only of order $k_s - k$ in N if k_s is equal to the largest of the k_r 's in the subscript and $k_1 + k_2 + \dots + k_N = k$. Going back to the expression of page 39, this shows that the term, $\lambda_{40} \lambda_{20}^2 \lambda_{11}^2$, for instance is of order -2 in N . The numerical coefficient is of order 2 in N , so that the result is a term of order -2 in N , since the general multiplier outside the brackets is of that order. Such a term is to be rejected and on similar grounds terms are thrown out until terms in $\lambda_{40} \lambda_{20}^4$ are left! But this is not all, for a computer with a little experience in this kind of computation would, on account of the special suitability of the D -operator for this purpose, never have computed but a few of the terms rejected, those on the

border line. Now the remaining terms are collected and there results,

$$\frac{49824 \lambda_{40} \lambda_{20}^4}{N^2}.$$

Another simplification in calculation arises here, for since 49824 is of order 2 in N , only terms of order zero in N need be saved in $\lambda_{40} \lambda_{20}^4$. Thus finally the only term of zero order in N arising from the whole expression on page 39 is $\frac{49824}{N^2} \lambda_4 \lambda_2^4$, and the only term of order

— 1 in N arising from this group of terms in $S_3(v_4)/\lambda_2^6$ is $\frac{49824}{N^2} \beta_4 \cdot 4!$.

Since this particular component term of $S_3(v_4)$ is as complicated as any that appear in it, and since for s and N both too the terms involving the higher orders of semi-invariants of $f(x)$ will drop out immediately in spite of the growing factorial multipliers involved, it is seen that under these assumptions it is easy to get an approximation to $S_3(v_4)/\lambda_2^6$ of order — 1 in N . Because the process is so reduced under these assumptions, either higher orders of approximation could be made, or smaller values of s and N , or both, could be used. And the great advantage of this process is that here the approximations are made sufficiently far along in the calculation for their effect on the final result to be watched and measured.

Frequency functions, $f(x)$, whose semi-invariants follow simple recursion laws offer interesting possibilities for investigation in sampling problems by this method. The most obvious is the case in which $f(x)$ is the normal curve of error. Here all the semi-invariants above the second are zero and every $S_r(v_m)$ and $S_{rs}(v_m, v_m)$ is at once reduced to at most one term. The study of this case alone is a matter of the greatest interest and importance. A second case of this kind is the one in which $f(x)$ is PEARSON'S Type III curve. If the equation be written, in standard units with the origin at the mean, in the form,

$$f(x) = \frac{b^b e^{-b}}{a^b \Gamma(b)} (a+x)^{b-1} e^{-\frac{b}{a}x} \quad (55)$$

then it is easily shown that,

$$\lambda_r = \frac{(r-1)! a^r}{b^{r-1}} \quad (56)$$

and

$$\beta_r = \frac{1}{r b^{r/2-1}} \quad (1)$$

(1) STEFFENSEN, J. F., *Matematisk Iagttagelseslaere*; G. E. C. Gads, Copenhagen, 1923; p. 60.

Since $b = \frac{4}{\alpha_3^2}$, here it can be seen just how much the β_r 's decrease in magnitude for any given degree of skewness, and then from the above a great deal can be learned about sampling in this very interesting case. A third case has already been mentioned, that in which it is assumed that $f(x)$ can be represented by means of the first three terms of a Type A series. Under the hypothesis of WICKSELL mentioned above another term may sometimes be added and $f(x)$ expressed,

$$f(x) = \varphi(x) + \beta_3 \varphi^{\text{III}}(x) + \beta_4 \varphi^{\text{IV}}(x) + \frac{\beta_3^2}{2} \varphi^{\text{VI}}(x), \quad (57)$$

(in which x is measured in standard units) a form first reached by Edgeworth ⁽¹⁾.

CHAPTER VII

The Semi-invariants of α_3 and α_4

It is the object of the present chapter to apply the method of semi-invariants and the results already obtained for the semi-invariants of the frequency functions of moments about the mean due to sampling to the task of finding better approximations than we have now to the values of the characteristics of the frequency functions of α_3 and α_4 due to sampling, where these quantities are defined by,

$$\alpha_3 = \frac{v_3}{v_2^{3/2}} ; \quad \alpha_4 = \frac{v_4}{v_2^2}$$

The fact that all but the first order semi-invariants are independent of the origin from which the variate is measured makes available a method which other investigators using moments alone could not employ.

A. For α_3 .

The semi-invariants, $b_1, b_2, b_3, \dots$ for α_3 are defined by,

$$e^{b_1 t + \frac{1}{2!} b_2 t^2 + \frac{1}{3!} b_3 t^3 + \dots} = \int_{-\infty}^{\infty} d\alpha_3 \varphi(\alpha_3) e^{i\alpha_3 t}, \quad (58)$$

⁽¹⁾ EDGEWORTH, F. Y., *The Law of Error*; « Cambridge Philosophical Transactions », Vol. XX, (1905) pp. 36-65; 113-132. See also WICKSELL, S. D. Loc. cit., p. 16.

with

$$\alpha_3 = \frac{v_3}{v_2^{3/2}},$$

and in which $\varphi(\alpha_3)$ is the probability function for α_3 .

Now I can write,

$$v_2 = v'_2 + \varepsilon_1, \quad (59)$$

in which v'_2 is the value of the second moment about the mean for the parent distribution, and ε_1 , then, is defined as the difference between the value of v_2 obtained from the sample and the true or « presumptive » value as THIELE called it. This is in reality, of course, only a change in origin for v_2 . Taking the mean of both sides of (59), I get,

$$\frac{N-1}{N} \lambda_2 = \lambda_2 + M_{\varepsilon_1}$$

whence it appears that the mean of the ε_1 's is $-\frac{1}{N} \lambda_2$ and is of the order -1 in N , a very important fact for what follows.

Writing,

$$\alpha_3 = \frac{v_3}{v_2^{3/2} \left(1 + \frac{\varepsilon_1}{v'_2}\right)^{3/2}} = \frac{v_3}{v_2'^{3/2} \left(1 + \frac{\varepsilon_1}{v'_2}\right)^{-3/2}}, \quad (60)$$

and expanding the binomial factor, I get an expression for α_3 (supposing there to be no difficulties over convergence, - a point to be discussed later) which enables me to substitute for $\varphi(\alpha_3)$ a probability function of which the semi-invariants are known, namely, $\theta(v_2, v_3)$, by which I designate the correlation function for v_2 and v_3 . That is, I write now,

$$\begin{aligned} & e^{b_1 t + \frac{1}{2!} b_2 t^2 + \frac{1}{3!} b_3 t^3 + \dots} \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} d v_2 d v_3 \theta(v_2, v_3) e^{\frac{v_3}{v_2'^{3/2} \left(1 + \frac{\varepsilon_1}{v'_2}\right)^{-3/2}} t} \end{aligned} \quad (61)$$

But still another shift of origin will be useful. I write

$$v_3 = v'_3 + \varepsilon_2,$$

in which v'_3 is the value of v_3 for the parent distribution and ε_2 the error, then, in v_3 due to sampling. Taking the mean of both sides of this equation,

$$\frac{(N-1)(N-2)}{N^2} \lambda_3 = \lambda_3 + M_{\varepsilon_2}$$

whence it appears that the mean of the ε_2 's is $\frac{-(3N-2)}{N^2} \lambda_3$ and is also of order -1 in N . Then (61) is written finally,

$$e^{b_1 t + \frac{1}{2!} b_2 t^2 + \frac{1}{3!} b_3 t^3 + \dots} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} d\nu_2 d\nu_3 \theta(\nu_2, \nu_3) e^{\frac{(\nu'_2 + \varepsilon_2)}{\nu'^{3/2}_2} \left(1 + \frac{\varepsilon_1}{\nu'_2}\right)^{-3/2} t} \quad (62)$$

On expanding and equating coefficients of t , I have,

$$b_1 = \frac{\lambda_3}{\nu'^{3/2}_2} \left(1 - \frac{3}{2} \frac{\bar{\nu}_{10}}{\nu'_2} + \frac{15}{8} \frac{\bar{\nu}_{20}}{\nu'^2_2} - \frac{35}{16} \frac{\bar{\nu}_{30}}{\nu'^3_2} + \dots \right) + \frac{1}{\nu'^{3/2}_2} \left(\bar{\nu}_{01} - \frac{3}{2} \frac{\bar{\nu}_{11}}{\nu'_2} + \frac{15}{8} \frac{\bar{\nu}_{21}}{\nu'^2_2} + \dots \right) \quad (63)$$

in which $\bar{\nu}_{rs}$ is the rs -th moment of $\theta(\nu_2, \nu_3)$,

$$b_2 = \frac{\lambda^2_3}{\nu'^3_2} \left(1 - 3 \frac{\bar{\nu}_{10}}{\nu'_2} + 6 \frac{\bar{\nu}^{20}_2}{\nu'^2_2} - 10 \frac{\bar{\nu}_{30}}{\nu'^3_2} + \dots \right) + 2 \frac{\lambda_3}{\nu'^3_2} \left(\bar{\nu}_{01} - 3 \frac{\bar{\nu}_{11}}{\nu'_2} + 6 \frac{\bar{\nu}_{21}}{\nu'^2_2} - 10 \frac{\bar{\nu}_{31}}{\nu'^3_2} + \dots \right) + \frac{1}{\nu'^3_2} \left(\bar{\nu}_{02} - 3 \frac{\bar{\nu}_{12}}{\nu'_2} + 6 \frac{\bar{\nu}_{22}}{\nu'^2_2} - 10 \frac{\bar{\nu}_{32}}{\nu'^3_2} + \dots \right) - b_1^2 \quad (64)$$

$$b_3 = \frac{\lambda^3_3}{\nu'^{9/2}_2} \left(1 - \frac{9}{2} \frac{\bar{\nu}_{10}}{\nu'_2} + \frac{99}{8} \frac{\bar{\nu}_{20}}{\nu'^2_2} - \frac{429}{16} \frac{\bar{\nu}_{30}}{\nu'^3_2} + \dots \right) + 3 \frac{\lambda^2_3}{\nu'^{9/2}_2} \left(\bar{\nu}_{01} - \frac{9}{2} \frac{\bar{\nu}_{11}}{\nu'_2} + \frac{99}{8} \frac{\bar{\nu}_{21}}{\nu'^2_2} - \frac{429}{16} \frac{\bar{\nu}_{31}}{\nu'^3_2} + \dots \right) + \frac{3\lambda_3}{\nu'^{9/2}_2} \left(\bar{\nu}_{02} - \frac{9}{2} \frac{\bar{\nu}_{12}}{\nu'_2} + \frac{99}{8} \frac{\bar{\nu}_{22}}{\nu'^2_2} - \frac{429}{16} \frac{\bar{\nu}_{32}}{\nu'^3_2} + \dots \right) + \frac{1}{\nu'^{9/2}_2} \left(\bar{\nu}_{03} - \frac{9}{2} \frac{\bar{\nu}_{13}}{\nu'_2} + \frac{99}{8} \frac{\bar{\nu}_{23}}{\nu'^2_2} - \frac{429}{16} \frac{\bar{\nu}_{33}}{\nu'^3_2} + \dots \right) - 3b_1 b_2 - b_1^3. \quad (65)$$

$$\begin{aligned}
b_4 = & \frac{\lambda_3^4}{v_2'^6} \left(\bar{1} - 6 \frac{\bar{v}_{10}}{v_2'} + 2\bar{1} \frac{\bar{v}_{20}}{v_2'^2} - 56 \frac{\bar{v}_{30}}{v_2'^3} + \dots \right) \\
& + 4 \frac{\lambda_3^3}{v_2'^6} \left(\bar{v}_{01} - 6 \frac{\bar{v}_{11}}{v_2'} + 2\bar{1} \frac{\bar{v}_{21}}{v_2'^2} - 56 \frac{\bar{v}_{31}}{v_2'^3} + \dots \right) \\
& + \frac{6 \lambda_3^2}{v_2'^6} \left(\bar{v}_{02} - 6 \frac{\bar{v}_{12}}{v_2'} + 2\bar{1} \frac{\bar{v}_{22}}{v_2'^2} - 56 \frac{\bar{v}_{32}}{v_2'^3} + \dots \right) \\
& + \frac{4 \lambda_3}{v_2'^6} \left(\bar{v}_{03} - 6 \frac{\bar{v}_{13}}{v_2'} + 2\bar{1} \frac{\bar{v}_{23}}{v_2'^2} - 56 \frac{\bar{v}_{33}}{v_2'^3} + \dots \right) \\
& + \frac{\bar{1}}{v_2'^6} \left(\bar{v}_{04} - 6 \frac{\bar{v}_{14}}{v_2'} + 2\bar{1} \frac{\bar{v}_{24}}{v_2'^2} - 56 \frac{\bar{v}_{34}}{v_2'^3} + \dots \right) \\
& - 4 \bar{b}_1 \bar{b}_3 - 3 \bar{b}_2^2 - 6 \bar{b}_2 \bar{b}_1^2 - \bar{b}_1^4.
\end{aligned} \tag{66}$$

To discuss the convergence of these developments let me begin with that for b_1 , looking first at the series,

$$\bar{1} - \frac{3}{2} \frac{\bar{v}_{10}}{v_2'} + \frac{15}{8} \frac{\bar{v}_{20}}{v_2'^2} - \frac{35}{16} \frac{\bar{v}_{30}}{v_2'^3} + \dots$$

The ratio of two consecutive numerical coefficients rapidly approaches unity. The moments $\frac{\bar{v}_{k0}}{v_2'^k}$ for low values of k at first rapidly decrease with increasing k , as may be seen from,

$$\begin{aligned}
\bar{v}_{10} &= S_{10} \\
\bar{v}_{20} &= S_{20} + S_{10}^2 \\
\bar{v}_{30} &= S_{30} + 3 S_{20} S_{10} + S_{10}^3
\end{aligned}$$

since $\frac{S_{k0}}{v_2'^k}$ is of order $1 - k$ in N for $k = 2, 3, 4, \dots$, and of order -1 in N for $k = 1$. But what happens for large values of k ? It is natural to inquire if $\frac{\bar{v}_{k0}}{v_2'^k}$ always decreases as k increases; if so convergence may be secured. Now a restriction upon the sampling which will certainly bring this about seems so unobjectionable that I shall take this way out. It is not a very severe restriction upon the distribution of sample v_2 's to rule out those which differ less than an arbitrarily small quantity from zero, or which shall then deviate from their mean by more than a certain fixed amount on the negative side. Now for even moderate values of N the distribution

of v_2 's has a very small skewness and it will not much increase the severity of the condition imposed to further restrict the positive deviations of the v_2 's considered from their mean to the same numerical quantity. Now the mean of v_2 's due to sampling is $\frac{N-1}{N}\lambda_2$ and my restriction means that only values of v_2 which lie *within*

$$\frac{N-1}{N}\lambda_2 - \frac{N-1}{N}\lambda_2 \quad \text{and} \quad \frac{N-1}{N}\lambda_2 + \frac{N-1}{N}v_2$$

are to be admitted to my distribution of sample v_2 's. Now let me repeat my question as to values of $\frac{\bar{v}_{k0}}{v_2'^k}$ for increasing k 's. I have,

$$\bar{v}_{k0} = \int_{-a\lambda_2}^{a\lambda_2} v_2^k P(v_2) dv_2$$

with $a < 1$, and $P(v_2)$, the frequency function for v_2 , which function is always positive and has a finite upper bound M . Further, then,

$$\bar{v}_{k0} \leq M \int_{-a\lambda_2}^{a\lambda_2} v_2^k dv_2 = \frac{2M a^{k+1} \lambda_2^{k+1}}{k+1}$$

and

$$\frac{\bar{v}_{k0}}{v_2'^k} \leq 2 \frac{M \lambda_2}{k+1} a^{k+1} \quad (v_2' = \lambda_2)$$

whence now $\frac{\bar{v}_{k0}}{v_2'^k}$ is decreasing with increasing k and it is apparent, moreover, that convergence has been secured.

Now noticing that the order of the moments with regard to v_3 is constant in the second series which occurs in the expression for b_1 , it is seen that the same argument applies in that case also and that thus that series will converge, too, under the same conditions. Moreover not only will the value for b_1 as expressed in (63) be convergent but all terms in it after the first few will be small. Besides the circumstance under the assumption made which contributes to this, as has just been explained, there is also another which helps with increasing orders of $\bar{v}_{kl}$, in which $k+l$ is said to be the order. In b_1 there are two terms of every order $k+l$ and their numerical coefficients are of opposite sign. The sum of these coefficients for terms of order

$k+l$ is easily shown by means of STIRLING'S formula to approach zero as $\frac{1}{\sqrt{k+l}}$, and since the $\bar{v}$'s of the same order of magnitude are also of the same order, here is another convergency factor.

Now on examining the expressions for b_2 , b_3 , and b_4 , the same arguments which were used in the convergency discussion for b_1 are found to apply here also in the case of each of these expressions. And besides, the remaining terms, the $-b_1^2$, in b_2 , the $-3b_1b_2 - b_1^4$ in b_3 , and the $-4b_1b_3 - 3b_2^2 - 6b_2b_1^2 - b_1^4$ in b_4 , are found when their previously obtained values are substituted to have the effect of cancelling out terms so as to leave in b_2 only terms of order -1 or less in N , in b_3 only terms of order -2 or less in N , and in b_4 only terms of order -3 or less in N .

In case the parent distribution is normal, $\sigma_{v_2} = \sqrt{2/N} \lambda_2$ and my restriction means that only sample v_2 's which fall within an interval of $\sqrt{N/2}$ of their own standard units from their mean are retained to form the distribution of sample v_2 's. If the parent distribution is not normal, the reciprocal of the more general expression for σ_{v_2} multiplied by λ_2 will be used instead of $\sqrt{N/2}$. In either case it is easy to judge of the severity of the restriction for any N .

Moreover the restriction will not have the effect of materially changing values of $\bar{v}_{kl}$'s of low orders. Thus the values obtained from the previous work for these quantities will be available and the error made will be on the safe side, that is, the values obtained will be too large rather than too small.

Indeed there seems to be no reason to suppose that the expressions given for the b 's do not converge under no restrictions whatever. However a proof of this seems to demand more information about the higher moments, $\bar{v}_{kl}$, than I have been able to get. And on the other hand, this restriction still leaves the results which are needed in practice. It ought, particularly, to be pointed out that all empirical tests made in order to check the results here found are bound to be carried out in such a way that the restrictions here made will be fulfilled.

I will now proceed to the determination of the b 's to the order of approximation which is secured by retaining all terms of order -3 and higher in N .

The necessary expressions for $\bar{v}_k$'s in terms of S 's are first written down. They are,

$$\begin{array}{l|l}
\overline{v}_{10} = S_{10} & \overline{v}_{11} = S_{11} + S_{10} S_{01} \\
\overline{v}_{01} = S_{01} & \overline{v}_{02} = S_{02} + S_{01}^2 \\
\overline{v}_{20} = S_{20} + S_{10}^2 & \\
\overline{v}_{30} = S_{30} + 3 S_{20} S_{10} + S_{10}^3 & \\
\overline{v}_{41} = S_{21} + S_{20} S_{01} + 2 S_{11} S_{10} + S_{10}^2 S_{01} & \\
\overline{v}_{12} = S_{12} + S_{02} S_{10} + 2 S_{11} S_{01} + S_{01}^2 S_{10} & \\
\overline{v}_{03} = S_{03} + 3 S_{02} S_{01} + S_{01}^3 & \\
\overline{v}_{40} = S_{40} + 4 S_{30} S_{10} + 3 S_{20}^2 S_{10} + 6 S_{20} S_{10}^2 + S_{10}^4 & \\
\overline{v}_{31} = S_{31} + S_{30} S_{01} + 3 S_{21} S_{10} + 3 S_{20} S_{11} + 3 S_{20} S_{10} S_{01} + 3 S_{11} S_{10}^2 + S_{10}^3 S_{01} & \\
\overline{v}_{22} = S_{22} + 2 S_{21} S_{01} + 2 S_{12} S_{10} + S_{20} S_{02} + 2 S_{11}^2 S_{10} + S_{20} S_{01}^2 + 4 S_{11} S_{10} S_{01} & \\
& + S_{02} S_{10}^2 + S_{10}^2 S_{01}^2 \\
\overline{v}_{13} = S_{13} + S_{03} S_{10} + 3 S_{12} S_{01} + 3 S_{02} S_{11} + 3 S_{02} S_{10} S_{01} + 3 S_{11} S_{01}^2 + S_{10} S_{01}^3 & \\
\overline{v}_{04} = S_{04} + 4 S_{03} S_{01} + 3 S_{02}^2 S_{01} + 6 S_{02} S_{01}^2 + S_{01}^4 & \\
\overline{v}_{30} = \dots + 15 S_{20}^2 S_{10} + \dots + 10 S_{30} S_{20} + \dots & \\
\overline{v}_{41} = \dots + 3 S_{20}^2 S_{01} + 12 S_{20} S_{11} S_{10} + \dots + 4 S_{30} S_{11} + 6 S_{21} S_{20} + \dots & \\
\overline{v}_{32} = \dots + 6 S_{20} S_{11} S_{01} + 3 S_{20} S_{02} S_{10} + 6 S_{11}^2 S_{10} + \dots + S_{30} S_{02} & \\
& + 6 S_{21} S_{11} + 3 S_{12} S_{20} + \dots \\
\overline{v}_{23} = \dots + 6 S_{02} S_{11} S_{10} + 3 S_{20} S_{02} S_{01} + 6 S_{11}^2 S_{01} + \dots + S_{03} S_{20} & \\
& + 6 S_{12} S_{11} + 3 S_{21} S_{02} + \dots \\
\overline{v}_{14} = \dots + 3 S_{02}^2 S_{10} + 12 S_{02} S_{11} S_{01} + \dots + 4 S_{03} S_{11} + 6 S_{12} S_{02} + \dots & \\
\overline{v}_{60} = \dots + 15 S_{20}^3 + \dots & \\
\overline{v}_{51} = \dots + 15 S_{20}^2 S_{11} + \dots & \\
\overline{v}_{42} = \dots + 3 S_{20}^2 S_{02} + 12 S_{20} S_{11}^2 + \dots & \\
\overline{v}_{33} = \dots + 9 S_{20} S_{11} S_{02} + 6 S_{11}^3 + \dots & \\
\overline{v}_{21} = \dots + 3 S_{20} S_{02}^2 + 12 S_{02} S_{11}^2 + \dots &
\end{array}$$

Substituting these values in (63) — (66) inclusive and reducing, I get, setting

$$\begin{aligned}
S_{kl} (v_m, v_n) / v_2^{\frac{k_m + l_n}{2}} &= g_{kl} (v_m, v_n), \\
b_1 = \alpha_1 \left(1 - \frac{3}{2} g_{10} + \frac{15}{8} g_{20} + \frac{15}{8} g_{10}^2 - \frac{35}{16} g_{30} - \frac{105}{16} g_{20} g_{10} - \frac{35}{16} g_{10}^3 \right. \\
&+ \frac{315}{128} g_{40} + \frac{315}{32} g_{30} g_{10} + \frac{945}{128} g_{20}^2 + \frac{945}{64} g_{20} g_{10}^2 - \frac{3465}{128} g_{30} g_{20} \\
&\left. - \frac{10395}{256} g_{20}^2 g_{10} + \frac{45045}{1024} g_{10}^3 \right) + \left(g_{01} - \frac{3}{2} g_{11} - \frac{3}{2} g_{10} g_{01} + \frac{15}{8} g_{21} \right.
\end{aligned}$$

$$\begin{aligned}
& + \frac{15}{8} g_{20} g_{01} + \frac{15}{4} g_{11} g_{10} + \frac{15}{8} g_{10}^2 g_{01} - \frac{35}{16} g_{31} - \frac{35}{16} g_{30} g_{01} - \frac{105}{16} g_{21} g_{10} \\
& - \frac{105}{16} g_{20} g_{11} - \frac{105}{16} g_{20} g_{10} g_{01} - \frac{105}{16} g_{11} g_{10}^2 \\
& + \frac{945}{128} g_{20}^2 g_{01} + \frac{945}{32} g_{20} g_{11} g_{10} + \frac{315}{32} g_{30} g_{11} + \frac{945}{64} g_{21} g_{20} \\
& - \frac{10395}{256} g_{20}^2 g_{11}
\end{aligned}$$

$$\begin{aligned}
b_2 = \alpha_2^3 & \left(\frac{9}{4} g_{20} - \frac{45}{8} g_{30} - \frac{45}{4} g_{20} g_{10} + \frac{645}{64} g_{40} + \frac{135}{4} g_{30} g_{10} + \frac{855}{32} g_{20}^2 \right. \\
& + \frac{135}{4} g_{20} g_{10}^2 - \frac{4725}{32} g_{30} g_{20} - \frac{5985}{32} g_{20}^2 g_{10} + \left. \frac{38955}{128} g_{20}^3 \right) \\
& + 2 \alpha_3 \left(-\frac{3}{2} g_{11} + \frac{33}{8} g_{21} + \frac{9}{4} g_{20} g_{01} + 6 g_{11} g_{10} - \frac{125}{16} g_{31} \right. \\
& - \frac{45}{8} g_{30} g_{01} - \frac{165}{8} g_{21} g_{20} - \frac{165}{8} g_{20} g_{11} - \frac{45}{4} g_{20} g_{10} g_{01} - 15 g_{11} g_{10}^2 \\
& + \frac{2295}{32} g_{21} g_{20} + \frac{375}{8} g_{30} g_{11} + \frac{855}{32} g_{20}^2 g_{01} + \frac{495}{4} g_{20} g_{11} g_{10} \\
& - \left. \frac{16065}{64} g_{20}^2 g_{11} \right) + \left(g_{02} - 3 g_{12} - 3 g_{02} g_{10} - 3 g_{11} g_{01} + 6 g_{22} \right. \\
& + \frac{33}{4} g_{21} g_{01} + 12 g_{12} g_{10} + 6 g_{20} g_{02} + \frac{39}{4} g_{11}^2 + \frac{9}{4} g_{20} g_{01}^2 \\
& + 12 g_{11} g_{10} g_{01} + 6 g_{10}^2 g_{02} - 10 g_{30} g_{02} - \frac{435}{8} g_{21} g_{11} - 30 g_{12} g_{20} \\
& - 30 g_{20} g_{02} g_{10} - \frac{165}{4} g_{20} g_{11} g_{01} - \frac{195}{4} g_{11}^2 g_{10} + 45 g_{20}^2 g_{02} \\
& + \left. \frac{2565}{16} g_{20} g_{11}^2 \right)
\end{aligned} \tag{68}$$

$$\begin{aligned}
b_3 = \alpha_3^3 & \left(-\frac{27}{8} g_{30} + \frac{405}{32} g_{40} + \frac{405}{16} g_{30} g_{10} + \frac{405}{16} g_{20}^2 - \frac{16605}{64} g_{30} g_{20} \right. \\
& - \frac{6885}{32} g_{20}^2 g_{10} + \left. \frac{88965}{128} g_{20}^3 \right) + 3 \alpha_2^3 \left(\frac{9}{4} g_{21} - 9 g_{31} - \frac{27}{8} g_{30} g_{01} \right. \\
& - \frac{117}{8} g_{21} g_{10} - 18 g_{20} g_{11} + \frac{3735}{32} g_{21} g_{20} + \frac{1215}{16} g_{30} g_{11} + \frac{405}{16} g_{20}^2 g_{01} \\
& + 135 g_{20} g_{11} g_{10} - \frac{17055}{32} g_{20}^2 g_{11} \left. \right) + 3 \alpha_3 \left(-\frac{3}{2} g_{12} + \frac{51}{8} g_{22} \right. \\
& + \frac{9}{2} g_{21} g_{01} + \frac{33}{4} g_{12} g_{10} + \frac{9}{2} g_{20} g_{02} + \frac{33}{4} g_{11}^2 - \frac{117}{8} g_{30} g_{02}
\end{aligned}$$

$$\begin{aligned}
& -\frac{663}{8} g_{21} g_{11} - \frac{717}{16} g_{12} g_{20} - \frac{117}{4} g_{20} g_{02} g_{10} - 36 g_{20} g_{11} g_{01} \\
& - \frac{429}{8} g_{11}^2 g_{10} + \frac{1395}{16} g_{20}^2 g_{02} + \frac{10215}{32} g_{20} g_{11}^2 + \left(g_{03} - \frac{9}{2} g_{13} \right. \\
& - \frac{9}{2} g_{03} g_{10} - \frac{9}{2} g_{12} g_{01} - 9 g_{02} g_{11} + \frac{99}{8} g_{03} g_{20} + \frac{243}{4} g_{12} g_{11} \\
& + \frac{63}{2} g_{21} g_{02} + \frac{27}{2} g_{20} g_{02} g_{01} + \frac{99}{2} g_{11} g_{02} g_{10} + \frac{99}{4} g_{11}^2 g_{01} \\
& \left. - \frac{1557}{8} g_{20} g_{11} g_{02} - \frac{909}{8} g_{11}^3 \right). \tag{68}
\end{aligned}$$

$$\begin{aligned}
b_4 = \alpha_3^4 & \left(\frac{81}{16} g_{40} - \frac{1215}{8} g_{20} g_{30} + \frac{4455}{8} g_{20}^3 \right) + 4 \alpha_3^3 \left(\frac{27}{8} g_{31} + \frac{81}{2} g_{30} g_{11} \right. \\
& + \frac{1053}{16} g_{21} g_{20} - 405 g_{20}^2 g_{11} \left. \right) + 6 \alpha_3^2 \left(-\frac{27}{4} g_{30} g_{02} + \frac{9}{4} g_{22} \right. \\
& - \frac{99}{4} g_{12} g_{20} + \frac{243}{4} g_{20}^2 g_{02} - \frac{171}{4} g_{21} g_{11} + \frac{1863}{8} g_{20} g_{11}^2 \left. \right) \\
& + 4 \alpha_3 \left(-\frac{3}{2} g_{13} + \frac{27}{4} g_{03} g_{20} + \frac{63}{2} g_{12} g_{11} + \frac{27}{2} g_{21} g_{02} - 84 g_{11}^3 \right. \\
& \left. - \frac{513}{4} g_{20} g_{11} g_{02} \right) + \left(g_{04} - 18 g_{03} g_{11} - 18 g_{12} g_{02} + 27 g_{20} g_{02}^2 + 126 g_{02} g_{11}^2 \right)
\end{aligned}$$

It is to be observed that in b_2 all terms of order zero in N cancel out, that in b_3 all terms of order -1 and higher in N cancel out, and that in b_4 all terms of order -2 and higher in N likewise vanish.

B. For α_4 .

The semi-invariants, $c_1, c_2, c_3, \dots$ of α_4 are defined by,

$$\begin{aligned}
e^{c_1 t + \frac{1}{2!} c_2 t^2 + \frac{1}{3!} c_3 t^3 + \dots} &= \int_{-\infty}^{\infty} d\alpha_4 \psi(\alpha_4) e^{\alpha_4 t}, \tag{69}
\end{aligned}$$

with $\alpha_4 = \frac{v_4}{v_2^2}$ and $\psi(\alpha_4)$ the probability function for α_4 .

Following the method of the last section, I write,

$$\alpha_4 = \frac{v_4' + \varepsilon_3}{v_2'^2 \left(1 + \frac{\varepsilon_1}{v_2'} \right)^2} = \frac{v_4' + \varepsilon_3}{v_2'^2} \left(1 + \frac{\varepsilon_1}{v_2'} \right)^{-2}, \tag{70}$$

in which both M_{ε_1} , and M_{ε_3} are of order $-\mathbf{1}$ in N . Writing now,

$$e^{c_1 t + \frac{1}{2!} c_2 t^2 + \frac{1}{3!} c_3 t^3 + \dots} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} d v_2 d v_4 \varphi(v_2, v_4) e^{\frac{v_4' + \varepsilon_3}{v_2'^2} \left(1 + \frac{\varepsilon_1}{v_2'}\right)^{-2} t} \quad (71)$$

I find,

$$c_1 = \frac{\bar{v}_4'}{v_2'^2} \left(\mathbf{1} - 2 \frac{\bar{v}_{10}}{v_2'} + 3 \frac{\bar{v}_{20}}{v_2'^2} - 4 \frac{\bar{v}_{30}}{v_2'^3} + 5 \frac{\bar{v}_{40}}{v_2'^4} - \dots \right) \quad (72)$$

$$+ \frac{\mathbf{1}}{v_2'^2} \left(\bar{v}_{01} - 2 \frac{\bar{v}_{11}}{v_2'} + 3 \frac{\bar{v}_{21}}{v_2'^2} - 4 \frac{\bar{v}_{31}}{v_2'^3} + 5 \frac{\bar{v}_{41}}{v_2'^4} - \dots \right)$$

$$c_2 = \frac{v_4'^2}{v_2'^4} \left(\mathbf{1} - 4 \frac{\bar{v}_{10}}{v_2'} + \mathbf{10} \frac{\bar{v}_{20}}{v_2'^2} - 20 \frac{\bar{v}_{30}}{v_2'^3} + \dots \right) \quad (73)$$

$$+ 2 \frac{v_4'}{v_2'^4} \left(\bar{v}_{01} - 4 \frac{\bar{v}_{11}}{v_2'} + \mathbf{10} \frac{\bar{v}_{21}}{v_2'^2} - 20 \frac{\bar{v}_{31}}{v_2'^3} + \dots \right)$$

$$+ \frac{\mathbf{1}}{v_2'^4} \left(\bar{v}_{02} - 4 \frac{\bar{v}_{12}}{v_2'} + \mathbf{10} \frac{\bar{v}_{22}}{v_2'^2} - 20 \frac{\bar{v}_{32}}{v_2'^3} + \dots \right)$$

$$- c_1^2$$

$$c_3 = \frac{v_4'^3}{v_2'^6} \left(\mathbf{1} - 6 \frac{\bar{v}_{10}}{v_2'} + 2\mathbf{1} \frac{\bar{v}_{20}}{v_2'^2} - 56 \frac{\bar{v}_{30}}{v_2'^3} + \dots \right) \quad (74)$$

$$+ 3 \frac{v_4'^2}{v_2'^6} \left(\bar{v}_{01} - 6 \frac{\bar{v}_{11}}{v_2'} + 2\mathbf{1} \frac{\bar{v}_{21}}{v_2'^2} - 56 \frac{\bar{v}_{31}}{v_2'^3} + \dots \right)$$

$$+ 3 \frac{v_4'}{v_2'^6} \left(\bar{v}_{02} - 6 \frac{\bar{v}_{12}}{v_2'} + 2\mathbf{1} \frac{\bar{v}_{22}}{v_2'^2} - 56 \frac{\bar{v}_{32}}{v_2'^3} + \dots \right)$$

$$+ \frac{\mathbf{1}}{v_2'^6} \left(\bar{v}_{03} - 6 \frac{\bar{v}_{13}}{v_2'} + 2\mathbf{1} \frac{\bar{v}_{23}}{v_2'^2} - 56 \frac{\bar{v}_{33}}{v_2'^3} + \dots \right)$$

$$- 3 c_1 c_2 - c_1^3.$$

$$c_4 = \frac{v_4'^4}{v_2'^8} \left(\mathbf{1} - 8 \frac{\bar{v}_{10}}{v_2'} + 36 \frac{\bar{v}_{20}}{v_2'^2} - \mathbf{120} \frac{\bar{v}_{30}}{v_2'^3} + \dots \right) \quad (75)$$

$$- 4 \frac{v_4'^3}{v_2'^8} \bar{v}_{01} - 8 \frac{\bar{v}_{11}}{v_2'} + 36 \frac{\bar{v}_{21}}{v_2'^2} - \mathbf{120} \frac{\bar{v}_{31}}{v_2'^3} + \dots$$

$$+ 6 \frac{v_4'^2}{v_2'^8} \left(\bar{v}_{02} - 8 \frac{\bar{v}_{12}}{v_2'} + 36 \frac{\bar{v}_{22}}{v_2'^2} - \mathbf{120} \frac{\bar{v}_{32}}{v_2'^3} + \dots \right)$$

$$\begin{aligned}
& + 4 \frac{v'_4}{v'^8_2} \left(\bar{v}_{03} - 8 \frac{\bar{v}_{13}}{v'_2} + 36 \frac{\bar{v}_{23}}{v'^2_2} - 120 \frac{\bar{v}_{33}}{v'^3_2} + \dots \right) \\
& + \frac{1}{v'^8_2} \left(\bar{v}_{04} - 8 \frac{\bar{v}_{14}}{v'_2} + 36 \frac{\bar{v}_{24}}{v'^2_2} - 120 \frac{\bar{v}_{34}}{v'^3_2} + \dots \right) \\
& - 4 c_1 c_3 - 3 c^2_2 - 6 c_2 c_1^2 - c_1^4.
\end{aligned} \tag{75}$$

The form of the c 's is essentially the same as for the b 's in the preceding section and the convergency discussion follows the same lines with the same kind of results.

The set of values of $\bar{v}_{kl}$'s in terms of the S 's used in the last section are also available here, with the difference, of course, that the $\bar{v}_{kl}$'s and the S_{ij} 's are now respectively the moments and the semi-invariants for the correlation function for v_2 and v_4 , instead of v_2 and v_3 as before. Making the substitutions, reducing, and saving terms of order -3 and over in N , I find, employing the g notation used for the b 's with the corresponding change in meaning,

$$\begin{aligned}
c_1 &= \alpha_4 (1 - 2 g_{10} + 3 g_{20} + 3 g^2_{10} - 4 g_{30} - 12 g_{20} g_{10} - 4 g^3_{10} + 5 g_{40} \\
&+ 20 g_{30} g_{10} + 15 g^2_{20} + 30 g_{20} g^2_{10} - 90 g^2_{20} g_{10} - 60 g_{30} g_{20} + 105 g^3_{20}) \\
&+ (g_{01} - 2 g_{11} - 2 g_{10} g_{01} + 3 g_{21} + 3 g_{20} g_{01} + 6 g_{11} g_{10} + 3 g^2_{10} g_{01} \\
&- 4 g_{31} - 4 g_{30} g_{01} - 12 g_{21} g_{10} - 12 g_{20} g_{11} - 12 g_{20} g_{10} g_{01} - 12 g_{11} g^2_{10} \\
&+ 15 g^2_{20} g_{01} + 60 g_{20} g_{11} g_{10} + 20 g_{30} g_{11} + 30 g_{21} g_{20} - 90 g^2_{20} g_{11}). \\
c_2 &= \alpha^2_4 (4 g_{20} - 12 g_{30} - 24 g_{20} g_{10} + 25 g_{40} + 84 g_{30} g_{10} + 66 g^2_{20} \\
&+ 84 g_{20} g^2_{10} - 528 g^2_{20} g_{10} - 416 g_{30} g_{20} + 960 g^3_{20}) + 2 \alpha_4 (-2 g_{11} \\
&+ 7 g_{21} + 4 g_{20} g_{01} + 10 g_{11} g_{10} - 16 g_{31} - 12 g_{30} g_{01} - 42 g_{21} g_{10} \\
&- 42 g_{20} g_{11} - 24 g_{20} g_{10} g_{01} - 30 g_{11} g^2_{10} + 66 g^2_{20} g_{01} + 294 g_{20} g_{11} g_{10} \\
&+ 112 g_{30} g_{11} + 171 g_{21} g_{20} - 684 g^2_{20} g_{11}) + (g_{02} - 4 g_{12} - 4 g_{02} g_{10} \\
&- 4 g_{11} g_{01} + 10 g_{22} + 14 g_{21} g_{01} + 20 g_{12} g_{10} + 10 g_{20} g_{02} + 16 g^2_{11} \\
&+ 4 g_{20} g^2_{01} + 10 g_{02} g^2_{10} + 20 g_{11} g_{10} g_{01} - 84 g_{20} g_{11} g_{01} - 60 g_{20} g_{02} g_{10} \\
&- 96 g^2_{11} g_{10} - 20 g_{30} g_{02} - 108 g_{21} g_{11} - 60 g_{12} g_{20} + 105 g^2_{20} g_{02} \\
&+ 372 g_{20} g^2_{11}). \\
c_3 &= \alpha^3_4 (-8 g_{30} + 36 g_{40} + 72 g_{30} g_{10} + 72 g^2_{20} - 720 g^2_{20} g_{10} - 864 g_{30} g_{20} \\
&+ 2664 g^3_{20}) + 3 \alpha^2_4 (4 g_{21} - 20 g_{31} - 8 g_{30} g_{01} - 32 g_{21} g_{10} - 40 g_{20} g_{11} \\
&+ 72 g^2_{20} g_{01} + 408 g_{20} g_{10} g_{11} + 204 g_{30} g_{11} + 312 g_{21} g_{20} - 1680 g^2_{20} g_{11}) \\
&+ 3 \alpha_4 (-2 g_{12} + 11 g_{22} + 8 g_{21} g_{01} + 14 g_{12} g_{10} + 8 g_{20} g_{02} + 14 g^2_{11} \\
&- 80 g_{20} g_{11} g_{01} - 64 g_{20} g_{02} g_{10} - 112 g^2_{11} g_{10} - 32 g_{30} g_{02} - 176 g_{21} g_{11} \\
&- 96 g_{12} g_{20} + 228 g^2_{20} g_{02} + 816 g_{20} g^2_{11}) + (g_{03} - 6 g_{13} - 6 g_{03} g_{10} \\
&- 6 g_{12} g_{01} - 12 g_{02} g_{11} + 84 g_{02} g_{11} g_{10} + 24 g_{20} g_{02} g_{01} + 42 g^2_{11} g_{01} \\
&+ 21 g_{06} g_{20} + 102 g_{12} g_{11} + 54 g_{21} g_{02} - 408 g_{20} g_{11} g_{02} - 232 g^3_{11}).
\end{aligned}$$

$$\begin{aligned}
c_4 = & \alpha_4^4 (16 g_{40} - 576 g_{30} g_{20} + 2496 g_{20}^3) + 4 \alpha_4^3 (-8 g_{31} + 120 g_{30} g_{11} \\
& + 192 g_{21} g_{20} - 1368 g_{20}^2 g_{11}) + 6 \alpha_4^2 (4 g_{22} - 16 g_{30} g_{02} - 96 g_{21} g_{11} \\
& - 56 g_{12} g_{20} + 176 g_{20}^2 g_{02} + 648 g_{20} g_{11}^2) + 4 \alpha_4 (-2 g_{13} + 12 g_{03} g_{20} \\
& + 54 g_{12} g_{11} + 24 g_{21} g_{02} - 288 g_{20} g_{11} g_{02} - 180 g_{11}^3) + (g_{04} - 24 g_{03} g_{11} \\
& - 24 g_{12} g_{02} + 48 g_{20} g_{02}^2 + 216 g_{11}^2 g_{02}).
\end{aligned}$$

C. *The Semi-invariants of $\sigma_x = \sqrt{v_2}$.*

The method developed for the semi-invariants of α_3 and α_4 can also be applied to the case of the standard deviation. It will be interesting in the case the parent distribution is taken to be normal to compare results with those already obtained by others so far as they go.

By definition the semi-invariants, $d_1, d_2, d_3, \dots$ of the standard deviation are given by,

$$\begin{aligned}
& e^{d_1 t + \frac{1}{2!} d_2 t^2 + \frac{1}{3!} d_3 t^3 + \dots} \\
& = \int_{-\infty}^{\infty} d\sigma_x f_1(\sigma_x) e^{\sigma_x t} \\
& = \int_{-\infty}^{\infty} d v_2 f_2(v_2) e^{\sqrt{v_2} t}
\end{aligned} \tag{77}$$

As before, I write,

$$v_2 = v'_2 + \varepsilon = \lambda_2 + \varepsilon$$

Then,

$$M_\varepsilon = -\frac{1}{N} \lambda_2,$$

and

$$\sqrt{v_2} = \sqrt{\lambda_2} \left(1 + \frac{\varepsilon}{\lambda_2}\right)^{\frac{1}{2}}$$

Rewriting, I have for the determination of the d 's,

$$\begin{aligned}
& e^{d_1 t + \frac{1}{2!} d_2 t^2 + \frac{1}{3!} d_3 t^3 + \dots} \\
& = \int_{-\infty}^{\infty} d v_2 f_2(v_2) e^{\lambda_2^{\frac{1}{2}} \left(1 + \frac{\varepsilon}{\lambda_2}\right)^{\frac{1}{2}} t}
\end{aligned} \tag{78}$$

This gives on expansion,

$$\begin{aligned}
 d_1 &= \sqrt{\lambda_2} \left(1 + \frac{1}{2} \frac{\bar{v}_1}{\lambda_2} - \frac{1}{8} \frac{\bar{v}_2}{\lambda_2^2} + \frac{1}{16} \frac{\bar{v}_3}{\lambda_2^3} - \frac{5}{128} \frac{\bar{v}_4}{\lambda_2^4} + \dots \right) \\
 d_2 &= \lambda_2 \left(1 + \frac{\bar{v}_1}{\lambda_2} \right) - d_1^2 \\
 d_3 &= \lambda_2^{3/2} \left(1 + \frac{3}{2} \frac{\bar{v}_1}{\lambda_2} + \frac{3}{8} \frac{\bar{v}_2}{\lambda_2^2} - \frac{1}{16} \frac{\bar{v}_3}{\lambda_2^3} + \frac{3}{128} \frac{\bar{v}_4}{\lambda_2^4} + \dots \right) \\
 &\quad - 3 d_1 d_2 - d_1^3. \\
 d_4 &= \lambda_2^2 (1 + 2 \bar{v}_1 + \bar{v}_2) - 4 d_1 d_3 - 3 d_2^2 - 6 d_2 d_1^2 - d_1^4.
 \end{aligned} \tag{79}$$

Once more I write down the expressions for the moments in terms of the semi-invariants.

$$\begin{aligned}
 \bar{v}_1 &= S_1 \\
 \bar{v}_2 &= S_2 + S_1^2 \\
 \bar{v}_3 &= S_3 + 3 S_2 S_1 + S_1^3 \\
 \bar{v}_4 &= S_4 + 4 S_3 S_1 + 3 S_2^2 + 6 S_2 S_1^2 + \dots \\
 \bar{v}_5 &= \dots + 10 S_3 S_2 + \dots + 15 S_2^3 S_1 + \dots \\
 \bar{v}_6 &= \dots + 15 S_3^2 + \dots
 \end{aligned} \tag{80}$$

(Only terms of order -3 or higher in N are saved). Then on substitution and saving terms of order -3 and higher in N , I have (writing a g_{r0} simply as a g_r),

$$\begin{aligned}
 d_1 &= \sigma_x \left(1 + \frac{1}{2} g_1 - \frac{1}{8} g_2 - \frac{1}{8} g_1^2 + \frac{1}{16} g_3 + \frac{3}{16} g_2 g_1 + \frac{1}{16} g_1^3 \right. \\
 &\quad \left. - \frac{5}{128} g_4 - \frac{5}{32} g_3 g_1 - \frac{15}{128} g_2^2 - \frac{15}{64} g_2 g_1^2 + \frac{35}{128} g_3 g_2 + \frac{105}{256} g_2^2 g_1 \right. \\
 &\quad \left. - \frac{315}{1024} g_3^2 \right) \\
 d_2 &= \frac{\lambda_2}{4} \left(g_2 - \frac{1}{2} g_3 - g_2 g_1 + \frac{5}{16} g_4 + g_3 g_1 + \frac{7}{8} g_2^2 + g_2 g_1^2 - \frac{17}{8} g_3 g_2 \right. \\
 &\quad \left. - \frac{21}{8} g_2^2 g_1 + \frac{75}{32} g_3^2 \right) \\
 d_3 &= \frac{\sigma_x^3}{8} \left(g_3 - \frac{3}{4} g_4 - \frac{3}{2} g_3 g_1 - \frac{3}{2} g_2^2 + \frac{39}{8} g_3 g_2 + \frac{15}{4} g_2^2 g_1 - \frac{83}{16} g_3^2 \right) \\
 d_4 &= \frac{\lambda_2^2}{16} (g_4 - 6 g_3 g_1 + 6 g_3^2).
 \end{aligned} \tag{81}$$

CHAPTER VIII

Results Thus Far Obtained

As one of the two concluding sections of this paper, I will now exhibit the results I have so far computed by the methods which have been developed herein.

For the characteristics of the correlation function for ν_2 and ν_3 , I have,

$$S_{10}(\nu_2, \nu_3) = S_1(\nu_2) = \frac{1}{N} (N-1) \lambda_2$$

$$S_{01}(\nu_2, \nu_3) = S_1(\nu_3) = \frac{1}{N^2} (N-1) (N-2) \lambda_3$$

$$S_{20}(\nu_2, \nu_3) = S_2(\nu_2) = \frac{1}{N^3} [(N-1)^2 \lambda_4 + 2 N (N-1) \lambda_2^2]$$

$$S_{11}(\nu_2, \nu_3) = \frac{1}{N^4} [(N-1)^2 (N-2) \lambda_5 + 6 N (N-1) (N-2) \lambda_3 \lambda_2]$$

$$S_{02}(\nu_2, \nu_3) = \frac{1}{N^5} [(N-1)^2 (N-2)^2 \lambda_6 + 9 N (N-1) (N-2)^2 (\lambda_4 \lambda_2 + \lambda_3^2) + 6 N^2 (N-1) (N-2) \lambda_3^3]$$

$$S_{30}(\nu_2, \nu_3) = \frac{1}{N^5} [(N-1)^3 \lambda_6 + 12 N (N-1)^2 \lambda_4 \lambda_2 + 4 N (N-1) (N-2) \lambda_3^2 + 8 N^2 (N-1) \lambda_3^3]$$

$$S_{21}(\nu_2, \nu_3) = \frac{1}{N^6} [(N-1)^3 (N-2) \lambda_7 + 16 N (N-1)^2 (N-2) \lambda_5 \lambda_2 + 12 N (N-1) (N-2) (2 N-3) \lambda_4 \lambda_3 + 48 N^2 (N-1) (N-2) \lambda_3 \lambda_2^2]$$

$$S_{12}(\nu_2, \nu_3) = \frac{1}{N^7} [(N-1)^3 (N-2)^2 \lambda_8 + 21 N (N-1)^2 (N-2)^2 \lambda_6 \lambda_2 + 6 N (N-1) (N-2)^2 (8 N-11) \lambda_5 \lambda_3 + 9 N (N-1) (N-2)^2 (3 N-5) \lambda_4^2 + 18 N^2 (N-1) (N-2) (6 N-11) \lambda_4 \lambda_2^2 + 18 N^2 (N-1) (N-2) (9 N-20) \lambda_3^2 \lambda_2 + 36 N^3 (N-1) (N-2) \lambda_3^4]$$

$$S_{03}(\nu_2, \nu_3) = \frac{1}{N^8} [N-1)^3 (N-2)^3 \lambda_9 + 27 N (N-1)^2 (N-2)^3 \lambda_7 \lambda_2 + 27 N (N-1) (N-2)^3 (3 N-4) \lambda_6 \lambda_3 + 27 N (N-1) (N-2)^3 (4 N-7) \lambda_5 \lambda_4 + 54 N^2 (N-1) (N-2)^2 (4 N-7) \lambda_5 \lambda_2^2]$$

$$\begin{aligned}
& + 162 N^2 (N-1) (N-2)^2 (5N-12) \lambda_4 \lambda_3 \lambda_2 \\
& + 36 N^2 (N-1) (N-2) (7N^2-30N+34) \lambda_3^3 \\
& + 108 N^3 (N-1) (N-2) (5N-12) \lambda_3 \lambda_2^3]
\end{aligned}$$

$$\begin{aligned}
S_{40}(\nu_2, \nu_3) = \frac{1}{N^7} [& N-1)^4 \lambda_8 + 24 N(N-1)^3 \lambda_6 \lambda_2 + 32 N(N-1)^2 (N-2) \lambda_5 \lambda_3 \\
& + 8 N(N-1) (4N^2-9N+6) \lambda_4^2 + 144 N^2 (N-1)^2 \lambda_4^2 \lambda_2^2 \\
& + 96 N^2 (N-1) (N-2) \lambda_3^2 \lambda_2 + 48 N^3 (N-1) \lambda_2^4].
\end{aligned}$$

$S_{31}(\nu_2, \nu_3)$, $S_{22}(\nu_1, \nu_3)$, $S_{13}(\nu_2, \nu_3)$ and $S_{04}(\nu_2, \nu_3)$. I have obtained in terms of the semi-invariants of $F(\delta_1, \dots, \delta_N)$; their expressions in this form are too long to reproduce here. Their complete reduction is a purely mechanical routine requiring no particular skill or ability. As I have explained, any term can be found independently of all the others, either in full or to any given degree of approximation. In particular, I have computed those end terms which do not vanish when the parent distribution is normal. In the case the parent distribution is not too far from normal (not too far to be successfully represented by a Type A series) these are the leading terms in the complete expressions for those S 's.

In the case, then, that the parent distribution is normal,

$$\begin{aligned}
S_{10}(\nu_2, \nu_3) &= \frac{1}{N} (N-1) \lambda_2 \\
S_{01}(\nu_2, \nu_3) &= 0 \\
S_{20}(\nu_2, \nu_3) &= \frac{2}{N^2} (N-1) \lambda_2^2 \\
S_{11}(\nu_2, \nu_3) &= 0 \\
S_{02}(\nu_2, \nu_3) &= \frac{6}{N^3} (N-1) (N-2) \lambda_2^3 \\
S_{30}(\nu_2, \nu_3) &= \frac{8}{N^3} (N-1) \lambda_2^3 \\
S_{21}(\nu_2, \nu_3) &= 0 \\
S_{12}(\nu_2, \nu_3) &= \frac{36 (N-1) (N-2) \lambda_2^4}{N^4} \\
S_{03}(\nu_2, \nu_3) &= 0 \\
S_{40}(\nu_2, \nu_3) &= \frac{48}{N^4} (N-1) \lambda_2^4 \\
S_{31}(\nu_2, \nu_3) &= 0
\end{aligned}$$

$$S_{22}(\nu_2, \nu_3) = \frac{288}{N^5} (N-1)(N-2)\lambda_2^5$$

$$S_{13}(\nu_2, \nu_3) = 0$$

$$S_{04}(\nu_2, \nu_3) = \frac{648}{N^6} (N-1)(N-2)(5N-12)\lambda_2^6.$$

For the characteristics of the correlation function for ν_2 and ν_4 I have,

$$S_{r0}(\nu_2, \nu_4) = S_{r0}(\nu_2, \nu_3)$$

$$S_{01}(\nu_2, \nu_4) = \frac{1}{N^3} [N-1](N^2-3N+3)\lambda_4 + 3N(N-1)^2\lambda_2^2]$$

$$S_{11}(\nu_2, \nu_4) = \frac{1}{N^5} [N-1]^2(N^2-3N+3)\lambda_6 \\ + 2N(N-1)(7N^2-18N+15)\lambda_4\lambda_2 \\ + 6N(N-1)(N-2)^2\lambda_3^2 + 12N^2(N-1)^2\lambda_3^2]$$

$$S_{02}(\nu_2, \nu_4) = \frac{1}{N^7} [(N-1)^2(N^2-3N+3)^2\lambda_8 \\ + 4N(N-1)(7N^4-39N^3+90N^2-99N+45)\lambda_6\lambda_2 \\ + 48N(N-1)(N-2)^2(N^2-3N+3)\lambda_5\lambda_3 \\ + 2N(N-1)(17N^4-111N^3+309N^2-405N+207)\lambda_4^2 \\ + 12N^2(N-1)(17N^3-71N^2+117N-69)\lambda_4\lambda_2^2 \\ + 72N^2(N-1)(N-2)^2(3N-5)\lambda_3^2\lambda_2 \\ + 24N^3(N-1)(4N^2-9N+6)\lambda_4^2]$$

$$S_{21}(\nu_2, \nu_4) = \frac{1}{N^7} [(N-1)^3(N^2-3N+3)\lambda_8 \\ + 2N(N-1)^2(13N^2-36N+33)\lambda_6\lambda_2 \\ + 8N(N-1)(N-2)(5N^2-15N+12)\lambda_5\lambda_3 \\ + 2N(N-1)(17N^3-71N^2+117N-69)\lambda_4^2 \\ + 16N^2(N-1)(11N^2-27N+21)\lambda_4\lambda_2^2 \\ + 24N^2(N-1)(N-2)(6N-11)\lambda_3^2\lambda_2 \\ + 72N^3(N-1)^2\lambda_2^4].$$

Of the missing semi-invariants of ν_2 and ν_4 up to the fourth order, I have S_{03} in terms of the semi-invariants of $F(\delta_1, \dots, \delta_N)$. For the others I have complete results only in the case the parent distribution is normal. In this case of the semi-invariants up to the fourth order follow. (Again $S_{r0}(\nu_2, \nu_4) = S_{r0}(\nu_2, \nu_3)$).

$$S_{01}(\nu_2, \nu_4) = \frac{3}{N^2} (N-1)^2\lambda_2^2$$

$$S_{11}(\nu_2, \nu_4) = \frac{12}{N^3} (N-1)^2\lambda_3^2$$

$$S_{02}(\nu_2, \nu_4) = \frac{24}{N^4} (N - 1) (4N^2 - 9N + 6) \lambda_2^4$$

$$S_{21}(\nu_2, \nu_4) = \frac{72}{N^4} (N - 1)^2 \lambda_2^4$$

$$S_{12}(\nu_2, \nu_4) = \frac{192}{N^5} (N - 1) (4N^2 - 9N + 6) \lambda_2^5$$

$$S_{03}(\nu_2, \nu_4) = \frac{864}{N^6} (N - 1) (11N^3 - 41N^2 + 59N - 31) \lambda_2^6$$

$$S_{31}(\nu_2, \nu_4) = \frac{576}{N^5} (N - 1)^2 \lambda_2^5$$

$$S_{22}(\nu_2, \nu_4) = \frac{1920}{N^6} (N - 1) (4N^2 - 9N + 6) \lambda_2^6$$

$$S_{13}(\nu_2, \nu_4) = \frac{10368}{N^7} (N - 1) (11N^3 - 41N^2 + 59N - 31) \lambda_2^7$$

$$S_{04}(\nu_2, \nu_4) = \frac{6912}{N^8} (N - 1) (272N^4 - 1467N^3 + 3363N^2 - 3717N + 1638) \lambda_2^8.$$

For the semi-invariants of α_3 , in case the parent distribution is normal, I have *to the order — 3 and higher in N*,

$$b_1 = 0$$

$$b_2 = \frac{6}{N} \left(1 - \frac{6}{N} + \frac{21}{N^2} \right)$$

$$b_3 = 0$$

$$b_4 = \frac{1296}{N^3}$$

For the semi-invariants of α_4 , in case the parent distribution is normal, I have *to the order — 3 and higher in N*,

$$c_1 = 3 \left(1 - \frac{2}{N} + \frac{2}{N^2} - \frac{5}{N^3} \right)$$

$$c_2 = \frac{24}{N} \left(1 - \frac{15}{N} + \frac{124}{N^2} \right)$$

$$c_3 = \frac{864}{N^2} \left(2 - \frac{11}{N} \right)$$

$$c_4 = \frac{684288}{N^3}$$

The semi-invariants of $\alpha_4 - 3$ which is $4!$ times the A_4 of the Type A series are, in the case parent distribution is normal, the same

as those given for α_4 immediately above with the exception that the first is,

$$c_1 - 3 = -\frac{2}{N} \left(1 - \frac{1}{N} + \frac{5}{2N^2} \right)$$

The semi-invariants of σ_x to the order -3 in N are,

$$d_1 = \sqrt{\lambda_2} \left(1 - \frac{3}{4N} - \frac{7}{32N^2} - \frac{9}{128N^3} \right)$$

$$d_2 = \frac{\lambda_2}{2N} \left(1 - \frac{1}{4N} - \frac{3}{8N^2} \right)$$

$$d_3 = \frac{\lambda_2^{3/2}}{4N^2} \left(1 + \frac{3}{4N} \right)$$

$$d_4 = 0.$$

which agree with previous results so far as they have been given ⁽¹⁾.

In this place it will also be of interest to point out that the simple law for moments about the mean for v_2 in samples from a normal parent distribution ⁽²⁾ appears in more simple form still when the semi-invariants of v_2 are given. Moreover, as can be seen by inspection of the results given above, the same kind of a simple law appears to hold for $S_{kl}(v_2, v_3)$ and $S_{kl}(v_2, v_4)$ in the case of a normal parent when l is held constant in either case and k is allowed to vary. The simplicity of what seems to be a general law was very tempting but so far I have had no luck in getting a general proof for it. Perhaps a really simple proof would furnish the key to a method of attack or a calculus which would greatly simplify the whole subject of this paper.

⁽¹⁾ See Editorial, *On distributions of the Standard Deviations of Small Samples*; Appendix No. 1 to Papers by STUDENT and R. A. FISHER; « Biometrika », Vol. X (1915), pp. 522-529.

See also ROMANOWSKY, V., *On the Moments of Standard Deviations and of Correlation Coefficients in Samples from Normal*; « Metron », Vol. V, No. 4 (1925), pp. 3-47.

⁽²⁾ See STUDENT, *On the Probable Error of a Mean*; « Biometrika » Vol. VI, (1908), pp. 1-25.

See also Editorial, « Biometrika », Loc. cit., and V. ROMANOWSKY Loc. cit.

CHAPTER IX

Some Empirically Obtained Results Compared with Those from the Theory

In an effort to get some empirical results to use as a check on the theory and for the purpose of illustrating the application of the theory, too, I made use of the facilities afforded by a HOLLERITH tabulating machine. The mechanical process was as follows.

First, from a table of areas under the normal frequency curve the following normal frequency distribution was made up:

f_x	x
1	0
2	1
9	2
28	3
66	4
121	5
175	6
197	7
175	8
121	9
66	10
28	11
9	12
2	13
1	14
1001	

Then 1001 cards were punched each with a value of x for x running from 0 to 14, the number of cards with each value of x being given by the above frequency table. And on each card, at the same time that the particular value of x was punched, there was also punched the square, the cube, and the fourth power of this value. Then the 1001 cards were thoroly mixed and remixed. From these was then selected at random a sample of 200. The tabulating machine counted the cards and gave for the 200, Σx , Σx^2 , Σx^3 , and Σx^4 , from which the various characteristics listed below were computed for the sample. The sample

was carefully reshuffled back into the pack and another sample was taken as before. The process was repeated until 400 samples were obtained. The mixing and the selecting were facilitated by separating the cards into sixteen piles into which small portions of samples were shuffled at random and from which samples were made up by taking from one to fifteen cards from one of the various piles quite at random until there were 200 or more in the hand.

Below are exhibited the distributions of the various sample characteristics that were found. For each characteristics are shown as computed and in comparison with those given by the theory for samples of 200 from a normal parent distribution which is infinite and from which an infinite number of samples is taken.

I. The Distribution of Sample Means.

Class intervals	f
6. 625-6. 674	0
6. 675-	2
6. 725-	12
6. 775-	20
6. 825-	22
6. 875-	46
6. 925-	57
6. 975-	66
7. 025-	55
7. 075-	50
7. 125-	40
7. 175-	15
7. 225-	11
7. 275-	3
7. 325-7. 374	1
	400

I list below the observed characteristics, that is, those computed from the distribution given just above, and those given by the known semi-invariants of sample means. THIELE gave ⁽¹⁾,

$$S_r(v'_1) = N^{1-r} \lambda_r.$$

⁽¹⁾ THIELE, T. N., Loc. cit., p. 42.

Then I have, for $N = 200$,

	Observed value	Predicted value
S_1	7.0055	7.0000
S_2	0.0149	0.0200
S_3	0.0001	0.0000
S_4	0.0001	0.0000

II. The Distribution of Sample v_2 's.

Class intervals	f
2.800-2.939	1
2.940-	0
3.080-	1
3.220-	7
3.360-	9
3.500-	25
3.640-	42
3.780-	48
3.920-	63
4.060-	60
4.200-	48
4.340-	40
4.480-	26
4.620-	14
4.760-	13
4.900-5.039	3
	400

For sample v_2 's, then I have for $N = 200$,

	Observed value	Predicted value
S_1	4.08	3.98
S_2	0.1301	0.1592
S_3	0.0017	0.0127
S_4	0.0031	0.0015

III. The Distribution of Sample v_3 's.

Class intervals	f
-5.200- -4.401	1
-4.400-	1
-3.600-	5
-2.800-	18
-2.000-	41

-1.200-	80
-0.400-	90
0.400-	81
1.200-	41
2.000-	33
2.800-	8
3.600-4.399	1
	400

Then for sample v_s 's, for $N = 200$,

	Observed values	Predicted values
S_1	0.1035	0.0000
S_2	1.92	1.89
S_3	-0.129	0.000
S_4	-0.036	1.614

IV. The Distribution of Sample v_4 's.

Class intervals	f
14.750-18.249	1
18.250-	1
21.750-	0
25.250-	1
28.750-	9
32.250-	17
35.750-	26
39.250-	37
42.750-	53
46.250-	55
49.750-	53
53.250-	44
56.750-	40
60.250-	23
63.750-	16
67.250-	12
70.750-	6
74.250-	3
77.750-	2
81.250-84.749	1
	400

For sample v_4 's, then, for $N = 200$

	Observed value	Predicted value
S_1	50.26	47.52
S_2	105.7	120.9
S_3	222.9	950.4
S_4	18433	14916.

V. The Distribution of Sample σ_x 's.

Class intervals	f
1.685-1.714	1
1.715-	0
1.745-	0
1.775-	4
1.805-	4
1.835-	4
1.865-	17
1.895-	29
1.925-	43
1.955-	39
1.985-	56
2.015-	54
2.045-	37
2.075-	44
2.105-	30
2.135-	15
2.165-	12
2.195-	9
2.225-2.254	2
	400

For sample σ_x 's, then, for $N = 200$,

	Observed value	Predicted value
d_1	2.0178	1.9925
d_2	0.0071	0.0099
d_3	0.0000	0.00005
d_4	0.0000	0.0000

VI. The Distribution of Sample σ_3 's.

Class intervals	f
-0.525-0.456	4
-0.455-	1
-0.385-	4
-0.315-	18
-0.245-	25
-0.175-	34
-0.105-	74
-0.035-	64
0.035-	59
0.105-	54
0.175	26
0.245	24
0.315-	9
0.385-	3
0.455-0.524	1
	400

For sample α_3 's, then, for $N = 200$,

	Observed value	Predicted value
b_1	0.0114	0.0000
b_2	0.028	0.029
b_3	-0.0005	0.0000
b_4	0.0001	0.0002

VII. The Distribution of Sample α_4 's.

Class intervals	f
1.330-1.469	1
1.470-	0
1.610-	0
1.750-	0
1.890-	1
2.030-	0
2.170-	4
2.310-	4
2.450-	27
2.590-	45
2.730-	64

2.870-	60
3.010-	68
3.150-	50
3.290-	38
3.430-	15
3.570-	10
3.710-	8
3.850-	4
3.990-4.139	1
	400

For sample α_4 's, for $N = 200$,

	Observed value	Predicted value
c_1	3.005	2.970
c_2	0.115	0.111
c_4	0.002	0.042
c_3	0.0147	0.0855

By finding in each case the standard deviation of the predicted characteristics, using now $N = 400$, I was able to conclude that the agreement between the empirical and the theoretical results was excellent in all cases except for the two higher characteristics of ν_4 's and α_4 's. Here the discrepancy seems rather serious but a little study of the frequency distributions of the samples of these two quantities suggested what appears to be the true explanation. The values of these two higher characteristics depend very largely upon the frequencies in the tails of the distributions. A little juggling of the values of these frequencies showed that but slight changes would be needed to effect the agreement desired or to make the disagreement twice as bad. Now the assumption on which the comparison between the empirical and the theoretical values is based is that the parent distribution from which samples are taken is infinite. As a matter of fact the parent population here has a total of 1001 members. In particular, the frequencies of large and small values of x in the parent population are actually quite small. The tails of the sample distributions are not by any means samples from an infinite population, not nearly enough so to give approximately the same results for the values of the two highest characteristics of the samples as if they were. For poor sampling in the tails will affect values of Σx^4 most and hence the values of ν_4 and α_4 most.

On the other hand since this is a case of sampling from a finite parent population an application of the theory developed for this kind of sampling is also relevant. In fact, Professor CARVER who has been working on this problem in sampling, tells me that his results lead one to expect just this sort of divergence between the results from sampling from infinite populations and from finite ones. Or really the well known results from the study of hypergeometric series as applied to the case of sampling without replacements are applicable here. It is especially striking that they at once predict what has happened of nearly all the second moments about the means of the sample distributions, namely, that the observed values are smaller than those that would be obtained from sampling with replacements, or what is the same thing, from sampling from an infinite population.

If a practical device were available for carrying on wholesale sampling with replacements we could get results that would be in all cases a more valid check on the theory.

CHAPTER X

Conclusions.

In the concluding section of this paper I will first summarize the contents.

1. The problem was stated and then a brief discussion of the general method of semi-invariants followed. The power of this method for the problem of sampling was made evident by showing how concisely it enables one get the semi-invariants of moments about a fixed point due to sampling.

2. The law which enables one to write semi-invariants in terms of moments and vice-versa was established ⁽¹⁾.

3. For the attack on the problem of getting the semi-invariants of moments about the mean due to sampling there was first developed the simple law for writing out the semi-invariants of $F(\delta_1 \dots, \delta_N)$, the correlation function of the deviations of sample values of x from their own mean.

⁽¹⁾ Since completing this paper I have talked with Dr RAGNAR FRISCH of the University of Oslo who tells me that he has recently published this law also in his doctor's thesis at the same university. I have not yet been able to see this paper myself, however.

4. By successive steps the general law for $S_{kl}(\nu_m, \nu_n)$ in terms of the moments of $F(\delta_1, \dots, \delta_N)$ was obtained and the possibility of writing down the general law for $S_{kl}(\nu_m, \nu_n)$ directly in terms of the semi-invariants of the parent distribution was shown.

5. The perfectly mechanical and mathematically simple process for calculating the values of $S_{kl}(\nu_m, \nu_n)$ in terms of the semi-invariants of the parent distribution was outlined and illustrated by actual calculations.

6. It was shown how certain terms of results may be obtained without reference to others and how results may be found to any degree of approximation with certainty and without calculation of terms afterwards to be discarded.

7. It was indicated how the semi-invariants of correlation functions of powers of moments are easily obtained once the semi-invariants of the moments themselves are found.

8. A method for obtaining the semi-invariants of σ_x , α_3 , and α_4 to a given degree of approximation was developed. By this means it is possible to give good results for these quantities in place of the very rough ones hitherto given. The process was carried out in detail in case the parent distribution is normal and in particular the agreement with the known results in the case of σ_x was noted.

9. A list of the results of my own computations of the semi-invariants of the moments about the mean and of two moments about the mean was given. In the case the parent distribution is normal complete results were given up to the characteristics of the fourth order for ν_2 , ν_3 , and ν_4 . This was partly for the purpose of comparison with results from some empirical sampling.

10. Finally the results of the experiment in sampling from a normal parent distribution were exhibited and the characteristics were compared with those obtained from the theory developed by this paper. The agreement was good in all cases but those in which the actual finiteness of the parent distribution makes itself felt.

Of course the investigation undertaken in this paper is only begun so far as actual algebraic and numerical results are concerned. But only a command of mechanical labor is needed to get all that may be required.

Finally, I must not fail to acknowledge my great indebtedness

and my gratitude to Professor SVEN D. WICKSELL of Lund, Sweden, under whose direction I began this paper and by the aid of whose suggestions and criticisms it was carried forward. Any merit it may have is due in a very large degree to his inspiration and help. And I must also express my obligation to Professors JAMES W. GLOVER and HARRY C. CARVER, under whom I continued my work after I returned to America and without whose cooperation and encouragement this investigation could not have been carried to its present stage.

Bibliography on Sampling from an Infinite Parent Distribution

« Biometrika » — Editorials:

1. *On the Distribution of the Standard Deviations of Small Samples*; Appendix to Papers by « STUDENT » and R. A. FISHER; Vol. X (1915), pp. 522-529.
2. *On the Probable Errors of Frequency Constants*; Vol. III (1903) pp. 273-281.
3. *On the Probable Errors of Frequency Constants*, Part II, Vol. IX (1913), pp. 1-10.
4. *On the Probable Errors of Frequency Constants*, Part III, Vol. XIII (1920), 113-132.

CHURCH, A. E. R.,

On the Moments of the Distribution of Squared Standard Deviations for Samples of N Drawn from an Indefinitely Large Population; « Biometrika » Vol XVII (1925), pp. 79-83.

CZUBER, EMANUEL,

Theorie der Beobachtungsfehler; B. G. Teubner, Leipzig (1891); p. 149.

FILON, L. N. G., and K. PEARSON,

On the Probable Errors of Frequency Constants and on the Influence of Random Selection on Variation and Correlation, « Philosophical Transactions of the Royal Society » Series A, Vol. CXCI (1898) p. 229.

FISHER, R. A.,

Frequency Distributions of the Values of the Correlation Coefficient in Samples from an Indefinitely Large Population; « Biometrika » Vol. X (1915), pp. 507-521.

HELMERT, F. R.,

Ueber die Wahrscheinlichkeit der Potenzsummen, usw.; « Zeitschrift für Mathematik und Physik » Vol. XXI (1876), p. 202.

HENDERSON, ROBERT,

Frequency Curves and Moments; « Transactions of the Actuarial Society of America » Vol. VIII, p. 30; also reprinted in the « Journal of the Institute of Actuaries », Vol. XLI (1907), pp. 429-442.

ISSERLIS, L.,

1. *On the Conditions under which the « Probable Errors » of Frequency Distributions have Real Significance*; « Proceedings of the Royal Society » Series A, Vol. XCII (1915), pp. 23-41.

2. *On the Value of a Mean as Calculated from a Sample*; « Journal of the Royal Statistical Society », Vol. LXXXI (1918), pp. 75-81.

JORGENSEN, N. R.,

Undersogelser over Frekvensflader og Korrelation; Arnold Busck, Copenhagen (1916), p. 122 ff.

NYDELL, STURE,

The Mean Errors of the Characteristics in Logarithmic Normal Distributions; « Skandinavisk Aktuarietidskrift » (1919), pp. 134-144.

PEARL, RAYMOND,

On Certain Points Concerning the Probable Errors of the Standard Deviation; « Biometrika », Vol. VI (1908), pp. 112-116.

PEARSON, KARL,

Further Contributions to the Theory of Small Samples; « Biometrika », Vol. XVII (1925), pp. 176-200.

PEARSON, KARL, and L. N. G. FILON,

See under FILON, L. N. G.,

ROMANOWSKY, V.,

On the Moments of Standard Deviations and of Correlation Coefficients in Samples from Normal; « Metron » Vol. V, No 4 (1925), pp. 3-47.

SHEPPARD, W. F.,

On the Application of the Theory of Error to Cases of Normal Distributions and Normal Correlations; « Philosophical Transactions of the Royal Society », Series A, Vol. CXCII (1898), pp. 101-167.

STEFFENSEN, J. F.,

Matematisk Iagttagelseslaere; G. E. C. Gode, Copenhagen (1923), p. 72 ff.

« STUDENT »,

1. *On the Distribution of the Means of Samples which are not drawn at Random*; « Biometrika », Vol. VII (1909), pp. 210-214.
2. *On the Probable Error of a Mean*; « Biometrika », Vol. VI, (1908), pp. 1-25.

TCHOUPROFF A. A.,

On the Mathematical Expectation of the Moments of Frequency Distributions; « Biometrika », Vol. XII (1918-1919), pp. 140-210; Vol. XIII (1921), pp. 283-308.

THIELE, T. N.,

1. *Elementaer Iagttagelseslaere*; Gylendalske Boghandel, Copenhagen (1897), p. 27 ff.
2. *Forelaesninger over Almendelig Iagttagelseslaere*; C. A. Reitzel, Copenhagen (1889), p. 58 ff.
3. *Theory of Observations*; C. & E. Layton, London (1903), p. 35 ff.

YASUKAWA, KAZUTARO

On the Means, Standard Deviations, Correlations, and Frequency Distributions of Variates; « Biometrika », Vol. XVII (1925), pp. 211-237.

Bibliography on Semi-invariants, etc.

CHARLIER, C. V. L.,

Contributions to the Mathematical Theory of Statistics, No. 5; « Arkiv för Matematik, Astronomi och Fysik », Band 9, No 25 (1914), or « Meddelanden från Lunds Astronomiska Observatorium », Series I, No. 57, p. 9.

EDGEWORTH, F. Y.,

The Law of Error; « Cambridge Philosophical Transactions », Volume XX (1905) pp. 36-65; 113-132.

FISCHER, ARNE,

The Mathematical Theory of Probabilities; Macmillan, New York (1922), p. 191 ff.

JORGENSEN, N. R.,

See under this name above, p. 8 ff.

MACMAHON, P. A.,

An Introduction to Combinatory Analysis; Cambridge University Press (1920), p. 24 ff.

STEFFENSEN, J. F.,

See under this name above, p. 21 ff.; p. 60.

THIELE, T. N.,

1. See under this name above, No. 1, p. 19 ff.
2. » » » » No. 2, p. 16 ff.
3. » » » » No. 3, p. 22 ff.

WICKSELL, S. D.,

The Correlation Function of Type A and the Regression of its Characteristics; « Kungliga Svenska Vetenskapsakademiens Handlingar », Band 58, No. 3, or « Meddelande från Lunds Astronomiska Observatorium » Series II, No. 17 (1917), p. 16 ff.
